PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07K 14/00 | A2

(11) International Publication Number: WO 98/55510

(43) International Publication Date: 10 December 1998 (10.12.98)

(21) International Application Number: PCT/US98/11735

(22) International Filing Date: 5 June 1998 (05.06.98)

(30) Priority Data:
08/870,143 5 June 1997 (05.06.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 08/870,143 (CIP)
Filed on 5 June 1997 (05.06.97)

(71) Applicant *(for all designated States except US)*: INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). GUEGLER, Karn, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). COCKS, Benjamin, Graeme [AU/US]; 4292 D. Wilke Way, Palo Alto, CA 94306 (US). SHAH, Purvi [IN/US]; 1608 Queen Charlotte Drive, Sunnyvale, CA 94087 (US).

(74) Agent: BILLINGS, Lucy, J.; Incyte Pharmaceuticals, Inc., 3147 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AT, AU, BR, CA, CH, CN, DE, DK, ES, FI, GB, IL, JP, KR, MX, NO, NZ, RU, SE, SG, US, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: CYCLIN RELATED PROTEINS

```
          9          18          27          36          45          54
CCC CAG AGG GTT ACA ATT CAA ACG CGG GCG GGC GGG CCC GCA GTC CTG CAG TTG

         63          72          81          90          99         108
CAG TCG TGT TCT CCG AGT TCC TGT CTC TCT GCC AAC GCC GCC CGG ATG GCT TCC
                                                          M   A   S

        117         126         135         144         153         162
CAA AAC CGC GAC CCA GCC GCC ACT AGC GTC GCC GCC GCC CGT AAA GGA GCT GAG
Q   N   R   D   P   A   A   T   S   V   A   A   A   R   K   G   A   E

        171         180         189         198         207         216
CCG AGC GGG GGC GCC GCC CGG GGT CCG GTG GGC AAA AGG CTA CAG CAG GAG CTG
P   S   G   G   A   A   R   G   P   V   G   K   R   L   Q   Q   E   L

        225         234         243         252         261         270
ATG ACC CTC ATG ATG TCT GGC GAT AAA GGG ATT TCT GCC TTC CCT GAA TCA GAC
M   T   L   M   M   S   G   D   K   G   I   S   A   F   P   E   S   D

        279         288         297         306         315         324
AAC CTT TTC AAA TGG GTA GGG ACC ATC CAT GGA GCA GCT GGA ACA GTA TAT GAA
N   L   F   K   W   V   G   T   I   H   G   A   A   G   T   V   Y   E
```

(57) Abstract

The invention provides human cyclin related proteins and polynucleotides which identify and encode HCRP. The invention also provides expression vectors, host cells, agonists, antibodies and antagonists. The invention also provides methods for treating disorders associated with expression of HCRP.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

CYCLIN RELATED PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of two cyclin related proteins and to the use of these sequences in the diagnosis, prevention, and treatment of cancer, immune disorders, and developmental disorders.

BACKGROUND OF THE INVENTION

Cell division is the fundamental process by which all living things grow and reproduce. In unicellular organisms such as yeast and bacteria, each cell division doubles the number of organisms, while in multicellular species many rounds of cell division are required to produce a new tissue or organ and to replace cells lost by wear or by programmed cell death. Details of the cell division cycle may vary, but the basic process consists of three principle events. The first event, interphase, involves preparations for cell division, replication of the DNA and production of essential proteins. In the second event, mitosis, the nuclear material is divided and separates to opposite sides of the cell. The final event, cytokinesis, is division and fission of the cell cytoplasm. The sequence and timing of these cell cycle events is under the control of the cell cycle control system which regulates the process at various check points. Over the past two decades, much research has been devoted to studying the structure and functions of various proteins that regulate these events.

The entry and exit of a cell from mitosis is regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins act by binding to and activating a group of cyclin-dependent protein kinases which then phosphorylate and activate selected proteins involved in the mitotic process. Several types of cyclins exist (Ciechanover, A. (1994) Cell 79:13-21). Two principle types are mitotic cyclin, or cyclin B, which controls entry of the cell into mitosis, and G1 cyclin, which controls events that drive the cell out of mitosis. Cyclins are characterized by a large region of shared homology that is approximately 180 amino acids in length and referred to as the "cyclin box" (Chapman, D.L. and Wolgemuth, D.J. (1993) Development 118:229-40). In addition, cyclins contain a sequence of amino acids in the N-terminal region of the molecule called the "destruction box"(Hunt, T. (1991) Nature 349:100-101). This sequence, which contains a conserved motif of RXXLXXIXN, is believed to be a recognition code that triggers ubiquitin-mediated degradation of cyclin B (see below).

Cyclins are degraded through the ubiquitin conjugation system (UCS), a major pathway

for the degradation of cellular proteins in eukaroytic cells and in some bacteria. The UCS mediates the elimination of abnormal proteins and regulates the half-lives of important regulatory proteins that control cellular processes such as gene transcription and cell cycle progression. The UCS is implicated in the degradation of mitotic cyclic kinases, oncoproteins, tumor suppressor genes such as p53, viral proteins, cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, supra).

The process of ubiquitin conjugation and protein degradation occurs in four principal steps (Jentsch, S. (1992) Annu. Rev. Genet. 26:179-207). First ubiquitin (Ub), a small, heat stable protein (76 amino acids) is activated by a ubiquitin-activating enzyme (E1) in an ATP dependent reaction which binds the C-terminus of Ub to the thiol group of an internal cysteine residue in E1. Second, activated Ub is transferred to one of several Ub-conjugating enzymes (E2). Different ubiquitin-dependent proteolytic pathways employ structurally similar, but distinct ubiquitin-conjugating enzymes that are associated with recognition subunits which direct them to proteins carrying a particular degradation signal. E2 then links the Ub molecule through its C-terminal glycine to an internal lysine (acceptor lysine) of a target protein. Additional Ub molecules may be added forming a multi-Ub chain structure. The ubiquinated protein is then recognized and degraded by proteasome, a large, multisubunit proteolytic enzyme complex, and Ub is released for reutilization.

The E2 (Ub-conjugating) enzymes are important for substrate specificity in different UCS pathways. All E2s have a conserved domain of approximately 16 kDa called the UBC domain that is at least 35% identical in all E2s and contains a centrally located cysteine residue required for ubiquitin-enzyme thiolester formation (Jentsch, supra). A well conserved proline-rich element is located N-terminal to the active cysteine residue. Structural variations beyond this conserved domain are used to classify the E2 enzymes. Class I E2s consist almost exclusively of the conserved UBC domain. Class II E2s have various unrelated C-terminal extensions that contribute to substrate specificity and cellular localization. Class III E2s have unique N-terminal extensions which are believed to be involved in enzyme regulation or substrate specificity.

Recently, a new type of E2 has been found that is selective for mitotic cyclins (Aris tarkhov, A.et al. (1996) Proc. Natl. Acad. Sci. 93:4294-99). Cyclin-specific E2, or E2-C, is characterized by the conserved UBC domain, an N-terminal extension of 30 amino acids not found in other E2s, and a unique sequence, TLLMSGD, adjacent to this extension. These characteristics together with the high affinity of E2-C for cyclin identify it as a new class of E2.

The discovery of new cyclin related proteins and the polynucleotides encoding them

satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention and treatment of cancer, immune disorders, and developmental disorders.

SUMMARY OF THE INVENTION

The invention features a substantially purified polypeptide, cyclin related protein HCRP-1 having the amino acid sequence shown in SEQ ID NO:1, or fragments thereof.

The invention further provides an isolated and substantially purified polynucleotide sequence encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or fragments thereof and a composition comprising said polynucleotide sequence. The invention also provides a polynucleotide sequence which hybridizes under stringent conditions to the polynucleotide sequence encoding the amino acid sequence SEQ ID NO:1, or fragments of said polynucleotide sequence. The invention further provides a polynucleotide sequence comprising the complement of the polynucleotide sequence encoding the amino acid sequence of SEQ ID NO:1, or fragments or variants of said polynucleotide sequence.

The invention also provides an isolated and purified sequence comprising SEQ ID NO.2 or variants thereof. In addition, the invention provides a polynucleotide sequence which hybridizes under stringent conditions to the polynucleotide sequence of SEQ ID NO:2. In another aspect the invention provides a composition comprising an isolated and purified polynucleotide sequence comprising the complement of SEQ ID NO:2, or fragments or variants thereof. The invention also provides a polynucleotide sequence comprising the complement of SEQ ID NO:2.

The present invention further provides an expression vector containing at least a fragment of any of the claimed polynucleotide sequences. In yet another aspect, the expression vector containing the polynucleotide sequence is contained within a host cell.

The invention also provides a method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment thereof, the method comprising the steps of: a) culturing the host cell containing an expression vector containing at least a fragment of the polynucleotide sequence encoding HCRP-1 under conditions suitable for the expression of the polypeptide; and b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified HCRP-1 having the amino acid sequence of SEQ ID NO:1 in conjunction with a suitable pharmaceutical carrier.

Still further, the invention provides a purified agonist which modulates the activity of the polypeptide of SEQ ID NO:1.

The invention also provides a method for treating or preventing cancer comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising purified HCRP-1.

The invention also provides a method for treating or preventing an immune disorder comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising purified HCRP-1.

The invention also provides a method for detecting a polynucleotide which encodes HCRP-1 in a biological sample comprising the steps of: a) hybridizing a polynucleotide sequence complementary to HCRP-1 (SEQ ID NO:1) to nucleic acid material of a biological sample, thereby forming a hybridization complex; and b) detecting the hybridization complex, wherein the presence of the complex correlates with the presence of a polynucleotide encoding HCRP-1 in the biological sample. In a preferred embodiment, prior to hybridization, the nucleic acid material of the biological sample is amplified by the polymerase chain reaction.

The invention also features a substantially purified polypeptide, cyclin related protein HCRP-2 having the amino acid sequence shown in SEQ ID NO:3, or fragments thereof.

The invention further provides an isolated and substantially purified polynucleotide sequence encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:3 or fragments thereof and a composition comprising said polynucleotide sequence. The invention also provides a polynucleotide sequence which hybridizes under stringent conditions to the polynucleotide sequence encoding the amino acid sequence SEQ ID NO:3, or fragments of said polynucleotide sequence. The invention further provides a polynucleotide sequence comprising the complement of the polynucleotide sequence encoding the amino acid sequence of SEQ ID NO:3, or fragments or variants of said polynucleotide sequence.

The invention also provides an isolated and purified sequence comprising SEQ ID NO:4 or variants thereof. In addition, the invention provides a polynucleotide sequence which hybridizes under stringent conditions to the polynucleotide sequence of SEQ ID NO:4. In another aspect the invention provides a composition comprising an isolated and purified polynucleotide sequence comprising the complement of SEQ ID NO:4, or fragments or variants thereof. The invention also provides a polynucleotide sequence comprising the complement of SEQ ID NO:4.

The present invention further provides an expression vector containing at least a fragment

of any of the claimed polynucleotide sequences. In yet another aspect, the expression vector containing the polynucleotide sequence is contained within a host cell.

The invention also provides a method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:3 or a fragment thereof, the method comprising the steps of: a) culturing the host cell containing an expression vector containing at least a fragment of the polynucleotide sequence encoding HCRP-2 under conditions suitable for the expression of the polypeptide; and b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified HCRP-2 having the amino acid sequence of SEQ ID NO:3 in conjunction with a suitable pharmaceutical carrier.

Still further, the invention provides a purified agonist which modulates the activity of the polypeptide of SEQ ID NO:3.

The invention also provides a purified antagonist which decreases the activity of a polypeptide of SEQ ID NO:3. In one aspect the invention provides a purified antibody which binds to a polypeptide comprising at least a fragment of the amino acid sequence of SEQ ID NO:3.

The invention also provides a method for treating or preventing a developmental disorder comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising purified HCRP-2.

The invention also provides a method for treating or preventing cancer comprising administering to a subject in need of such treatment an effective amount of a purified antagonist of HCRP-2.

The invention also provides a method for detecting a polynucleotide which encodes HCRP-2 in a biological sample comprising the steps of: a) hybridizing a polynucleotide sequence complementary to HCRP-2 (SEQ ID NO:3) to nucleic acid material of a biological sample, thereby forming a hybridization complex; and b) detecting the hybridization complex, wherein the presence of the complex correlates with the presence of a polynucleotide encoding HCRP-1 in the biological sample. In a preferred embodiment, prior to hybridization, the nucleic acid material of the biological sample is amplified by the polymerase chain reaction.

BRIEF DESCRIPTION OF THE FIGURES

Figures 1A, 1B, and 1C show the amino acid sequence (SEQ ID NO:1) and nucleic acid sequence (SEQ ID NO:2) of HCRP-1. The alignment was produced using MacDNASIS PRO™

software (Hitachi Software Engineering Co. Ltd. San Bruno, CA).

Figures 2A, 2B, 2C, 2D, 2E, and 2F show the amino acid sequence (SEQ ID NO:3) and nucleic acid sequence (SEQ ID NO:4) of HCRP-2. The alignment was produced using MacDNASIS PRO™ software.

Figure 3 shows the amino acid sequence alignments among HCRP-1 (SEQ ID NO:1), the cyclin-specific ubiquitin carrier protein E2-C from atlantic clam (GI 1493838; SEQ ID NO:5) and the ubiquitin carrier protein E2-18KD from yeast (GI 171866; SEQ ID NO:6), produced using the multisequence alignment program of DNASTAR™ software (DNASTAR Inc, Madison WI).

Figures 4A and 4B show the amino acid sequence alignments among HCRP-2 (SEQ ID NO:3), and cyclin B from mouse (GI 50613; SEQ ID NO:7) and human (GI 105763; SEQ ID NO:8), produced using the multisequence alignment program of DNASTAR™ software.

Figures 5A and 5B show the hydrophobicity plots for HCRP-1, SEQ ID NO: 1 and cyclin-specific ubiquitin carrier protein E2-C from atlantic clam (SEQ ID NO:5), respectively; the positive X axis reflects amino acid position, and the negative Y axis, hydrophobicity (MacDNASIS PRO software).

Figures 6A and 6B show the hydrophobicity plots for HCRP-2, SEQ ID NO: 3 and cyclin B from mouse (SEQ ID NO:7), respectively ; the positive X axis reflects amino acid position, and the negative Y axis, hydrophobicity (MacDNASIS PRO software).

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a host cell" includes a plurality of such host cells, reference to the "antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same

meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

HCRP, as used herein, refers to the amino acid sequences of substantially purified HCRP obtained from any species, particularly mammalian, including bovine, ovine, porcine, murine, equine, and preferably human, from any source whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist", as used herein, refers to a molecule which, when bound to HCRP, increases or prolongs the duration of the effect of HCRP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of HCRP.

An "allele" or "allelic sequence", as used herein, is an alternative form of the gene encoding HCRP. Alleles may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding HCRP as used herein include those with deletions, insertions, or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent HCRP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding HCRP, and improper or unexpected hybridization to alleles, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding HCRP. The encoded protein may also be "altered" and contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally

equivalent HCRP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological or immunological activity of HCRP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid; positively charged amino acids may include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine, glycine and alanine, asparagine and glutamine, serine and threonine, and phenylalanine and tyrosine.

"Amino acid sequence" as used herein refers to an oligopeptide, peptide, polypeptide, or protein sequence, and fragment thereof, and to naturally occurring or synthetic molecules. Fragments of HCRP are preferably about 5 to about 15 amino acids in length and retain the biological activity or the immunological activity of HCRP. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, amino acid sequence, and like terms, are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule.

"Amplification" as used herein refers to the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction (PCR) technologies well known in the art (Dieffenbach, C.W. and G.S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY).

The term "antagonist" as used herein, refers to a molecule which, when bound to HCRP, decreases the amount or the duration of the effect of the biological or immunological activity of HCRP. Antagonists may include proteins. nucleic acids, carbohydrates, or any other molecules which decrease the effect of HCRP.

As used herein, the term "antibody" refers to intact molecules as well as fragments thereof, such as Fa, $F(ab')_2$, and Fv, which are capable of binding the epitopic determinant. Antibodies that bind HCRP polypeptides can be prepared using intact polypeptides or fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal can be derived from the translation of RNA or synthesized chemically and can be conjugated to a carrier protein, if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin and thyroglobulin, keyhole limpet hemocyanin. The coupled peptide is then used to immunize the animal (e.g., a mouse, a rat, or a rabbit).

The term "antigenic determinant", as used herein, refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or fragment of a

protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to a given region or three-dimensional structure on the protein; these regions or structures are referred to as antigenic determinants. An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense", as used herein, refers to any composition containing nucleotide sequences which are complementary to a specific DNA or RNA sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. Antisense molecules include peptide nucleic acids and may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and block either transcription or translation. The designation "negative" is sometimes used in reference to the antisense strand, and "positive" is sometimes used in reference to the sense strand.

The term "biologically active", as used herein, refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic HCRP, or any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity", as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base-pairing. For example, the sequence "A-G-T" binds to the complementary sequence "T-C-A". Complementarity between two single-stranded molecules may be "partial", in which only some of the nucleic acids bind, or it may be complete when total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands and in the design and use of PNA molecules.

A "composition comprising a given polynucleotide sequence" as used herein refers broadly to any composition containing the given polynucleotide sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding HCRP (SEQ ID NO:1 or SEQ ID NO:3) or fragments thereof (e.g., SEQ ID NO:2 of SEQ ID NO:4 and fragments thereof) may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as

a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., SDS) and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus", as used herein, refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, has been extended using XL-PCR™ (Perkin Elmer, Norwalk, CT) in the 5' and/or the 3' direction and resequenced, or has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly (e.g., GELVIEW™ Fragment Assembly system, GCG, Madison, WI). Some sequences have been both extended and assembled to produce the consensus sequence .

The term "correlates with expression of a polynucleotide", as used herein, indicates that the detection of the presence of ribonucleic acid that is similar to SEQ ID NO:2 or SEQ ID NO:4 by northern analysis is indicative of the presence of mRNA encoding HCRP in a sample and thereby correlates with expression of the transcript from the polynucleotide encoding the protein.

A "deletion", as used herein, refers to a change in the amino acid or nucleotide sequence and results in the absence of one or more amino acid residues or nucleotides.

The term "derivative", as used herein, refers to the chemical modification of a nucleic acid encoding or complementary to HCRP or the encoded HCRP. Such modifications include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative encodes a polypeptide which retains the biological or immunological function of the natural molecule. A derivative polypeptide is one which is modified by glycosylation, pegylation, or any similar process which retains the biological or immunological function of the polypeptide from which it was derived.

The term "homology", as used herein, refers to a degree of complementarity. There may be partial homology or complete homology (i.e., identity). A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or hybridization probe will compete for and inhibit the binding of a completely homologous sequence to the target sequence under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific

binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% identity). In the absence of non-specific binding, the probe will not hybridize to the second non-complementary target sequence.

Human artificial chromosomes (HACs) are linear microchromosomes which may contain DNA sequences of 10K to 10M in size and contain all of the elements required for stable mitotic chromosome segregation and maintenance (Harrington, J.J. et al. (1997) Nat Genet. 15:345-355).

The term "humanized antibody", as used herein, refers to antibody molecules in which amino acids have been replaced in the non-antigen binding regions in order to more closely resemble a human antibody, while still retaining the original binding ability.

The term "hybridization", as used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

The term "hybridization complex", as used herein, refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary G and C bases and between complementary A and T bases; these hydrogen bonds may be further stabilized by base stacking interactions. The two complementary nucleic acid sequences hydrogen bond in an antiparallel configuration. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

An "insertion" or "addition", as used herein, refers to a change in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, as compared to the naturally occurring molecule.

"Microarray" refers to a high-density array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support.

The term "modulate", as used herein, refers to a change in the activity of HCRP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional or immunological properties of HCRP.

"Nucleic acid sequence" as used herein refers to an oligonucleotide, nucleotide, or polynucleotide, and fragments thereof, and to DNA or RNA of genomic or synthetic origin which may be single- or double-stranded, and represent the sense or antisense strand. "Fragments" are those nucleic acid sequences which are greater than 60 nucleotides than in length, and most

preferably includes fragments that are at least 100 nucleotides or at least 1000 nucleotides, and at least 10.000 nucleotides in length.

The term "oligonucleotide" refers to a nucleic acid sequence of at least about 6 nucleotides to about 60 nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20 to 25 nucleotides, which can be used in PCR amplification or hybridization assays. As used herein, oligonucleotide is substantially equivalent to the terms "amplimers","primers", "oligomers", and "probes", as commonly defined in the art.

"Peptide nucleic acid", PNA as used herein, refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least five nucleotides in length linked to a peptide backbone of amino acid residues which ends in lysine. The terminal lysine confers solubility to the composition. PNAs may be pegylated to extend their lifespan in the cell where they preferentially bind complementary single stranded DNA and RNA and stop transcript elongation (Nielsen, P.E. et al. (1993) Anticancer Drug Des. 8:53-63).

The term "portion", as used herein, with regard to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size from five amino acid residues to the entire amino acid sequence minus one amino acid. Thus, a protein "comprising at least a portion of the amino acid sequence of SEQ ID NO:1 or SEQ ID NO:3" encompasses the full-length HCRP and fragments thereof.

The term "sample", as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acid encoding HCRP, or fragments thereof, or HCRP itself may comprise a bodily fluid, extract from a cell, chromosome, organelle, or membrane isolated from a cell, a cell, genomic DNA, RNA, or cDNA(in solution or bound to a solid support, a tissue, a tissue print, and the like.

The terms "specific binding" or "specifically binding", as used herein, refers to that interaction between a protein or peptide and an agonist, an antibody and an antagonist. The interaction is dependent upon the presence of a particular structure (i.e., the antigenic determinant or epitope) of the protein recognized by the binding molecule. For example, if an antibody is specific for epitope "A", the presence of a protein containing epitope A (or free, unlabeled A) in a reaction containing labeled "A" and the antibody will reduce the amount of labeled A bound to the antibody.

The terms "stringent conditions"or "stringency", as used herein, refer to the conditions for hybridization as defined by the nucleic acid, salt, and temperature. These conditions are well known in the art and may be altered in order to identify or detect identical or related

polynucleotide sequences. Numerous equivalent conditions comprising either low or high stringency depend on factors such as the length and nature of the sequence (DNA, RNA, base composition), nature of the target (DNA, RNA, base composition), milieu (in solution or immobilized on a solid substrate), concentration of salts and other components (e.g., formamide, dextran sulfate and/or polyethylene glycol), and temperature of the reactions (within a range from about 5°C below the melting temperature of the probe to about 20°C to 25°C below the melting temperature). One or more factors be may be varied to generate conditions of either low or high stringency different from, but equivalent to, the above listed conditions.

The term "substantially purified", as used herein, refers to nucleic or amino acid sequences that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably 90% free from other components with which they are naturally associated.

A "substitution", as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Transformation", as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. It may occur under natural or artificial conditions using various methods well known in the art. Transformation may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. Such "transformed" cells include stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome. They also include cells which transiently express the inserted DNA or RNA for limited periods of time.

A "variant" of HCRP, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties, e.g., replacement of leucine with isoleucine. More rarely, a variant may have "nonconservative" changes, e.g., replacement of a glycine with a tryptophan. Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR software.

THE INVENTION

The invention is based on the discovery of new human cyclin related proteins (hereinafter referred to as "HCRP"), the polynucleotides encoding HCRP, and the use of these compositions for the diagnosis, prevention, or treatment of cancer, immune disorders, and developmental disorders.

In one embodiment, the invention encompasses a purified related protein (HCRP-1) comprising the amino acid sequence of SEQ ID NO:1, as shown in Figure 1. HCRP-1 is 179 amino acids in length and has a potential myrsitoylation site at residue G19. A potential casein kinase 2 phosphorylation site is found at T72, and a potential protein kinase C phosphorylation site is found at T95. A leucine zipper motif is found beginning at L118 and consisting of leucine residues repeated at regular intervals at L118, L125, L132, and L139. As shown in Figure 3, HCRP-1 has chemical and structural homology with the cyclin-specific ubiquitin carrier protein E2-C from atlantic clam (GI 1493838; SEQ ID NO:5) and the ubiquitin carrier protein E2-18KD from yeast (GI 171866; SEQ ID NO:6). In particular, HCRP-1 shares 61% and 53% identity with E2-C and E2-18KD, respectively. The active site cysteine residue of Ub-conjugating enzymes is found in HCRP-1 at C114 and is conserved in both E2-C and E2-18KD. The UBC domain surrounding this active site extends from approximately Y103 to L118 and is also highly conserved in the other E2s. A proline rich region is found in HCRP-1 on the N-terminal side of the active cysteine residue. Proline residues located at P86, P90, P94, P101, and P105 in HCRP-1 are conserved in the other E2s. HCRP-1 contains a unique N-terminal extension characteristic of class III E2s which spans residues A14 through A27. As illustrated by Figs. 5A and 5B, HCRP-1 and E2-C have rather similar hydrophobicity plots. Northern analysis shows the expression of this sequence in various libraries, at least 51% of which are immortalized or cancerous and at least 24% of which involve inflammation and the immune response. Of particular note is the expression of HCRP-1 in carcinomas of the bladder and pancreas, in metastatic tumors in the brain and lung, and in inflammed tissues associated with rehumatoid arthritis.

Nucleic acids encoding the HCRP-1 of the present invention were first identified in Incyte Clone 1308478 from the fetal colon cDNA library (COLNFET02) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:2, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 993201/ COLNNOT11, 1308478/ COLNFET02, and 2342963/ TESTTUT02.

In another embodiment, the invention encompasses a purified cyclin related protein

(HCRP-2) comprising the amino acid sequence of SEQ ID NO:3, as shown in Figure 2. HCRP-2 is 398 amino acids in length and has a typical "cyclin box", signature sequence of 180 amino acids extending from M165 through S313. A putative "destruction box" sequence is found beginning at R32 (RTVLEEIGN). As shown in Figure 4, HCRP-2 has chemical and structural homology with cyclin B from mouse (GI 50613; SEQ ID NO:7) and human (GI 105763; SEQ ID NO:8). In particular, HCRP-2 shares 89% and 54% identity with mousse and human cyclin B, respectively. The region of the "cyclin box" in HCRP-2 is well conserved in both mouse and human cyclin B. The "destruction box" for ubiquitin-mediated degradation is also well conserved in the mouse and human cyclins. As shown in Figure 6A and 6B, HCRP-2 and the mouse cyclin B have rather similar hydrophobicity plots. Northern analysis shows the expression of this sequence in various libraries, at least 70% of which are immortalized or cancerous and 20% of which involve fetal tissues.

Nucleic acids encoding the HCRP-2 of the present invention were first identified in Incyte Clone 7755 from the human mast cell line cDNA library (HMC1NOT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:4, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 7755/ HMC1NOT01, 256850 and 484361/ HNT2RAT01, and 441496/ MPHGNOT03.

The invention also encompasses HCRP variants. A preferred HCRP variant is one having at least 80%, and more preferably 90%, amino acid sequence identity to the HCRP amino acid sequence (SEQ ID NO:1 or SEQ ID NO:3). A most preferred HCRP variant is one having at least 95% amino acid sequence identity to SEQ ID NO:1 or SEQ ID NO:3.

The invention also encompasses polynucleotides which encode HCRP. Accordingly, any nucleic acid sequence which encodes the amino acid sequence of HCRP can be used to produce recombinant molecules which express HCRP. In a particular embodiment, the invention encompasses the polynucleotide comprising the nucleic acid sequence of SEQ ID NO:2 or SEQ ID NO:4 as shown in Figure 1 or Figure 3, respectively.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding HCRP, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequence of naturally occurring HCRP, and all such variations are to be considered as

being specifically disclosed.

Although nucleotide sequences which encode HCRP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring HCRP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding HCRP or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding HCRP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences, or fragments thereof, which encode HCRP and its derivatives, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding HCRP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed nucleotide sequences, and in particular, those shown in SEQ ID NO:2 or SEQ ID NO:4, under various conditions of stringency as taught in Wahl, G.M. and S.L. Berger (1987; Methods Enzymol. 152:399-407) and Kimmel, A.R. (1987; Methods Enzymol. 152:507-511).

Methods for DNA sequencing which are well known and generally available in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp, Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE Amplification System marketed by Gibco/BRL (Gaithersburg, MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

The nucleic acid sequences encoding HCRP may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be

employed, "restriction-site" PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). In particular, genomic DNA is first amplified in the presence of primer to a linker sequence and a primer specific to the known region. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). The primers may be designed using commercially available software such as OLIGO 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, MN), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72° C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may also be used to place an engineered double-stranded sequence into an unknown fragment of the DNA molecule before performing PCR.

Another method which may be used to retrieve unknown sequences is that of Parker, J.D. et al. (1991; Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions. When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In

particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and detection of the emitted wavelengths by a charge coupled devise camera. Output/light intensity may be converted to electrical signal using appropriate software (e.g. Genotyper™ and Sequence Navigator™, Perkin Elmer) and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode HCRP may be used in recombinant DNA molecules to direct expression of HCRP, fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced, and these sequences may be used to clone and express HCRP.

As will be understood by those of skill in the art, it may be advantageous to produce HCRP-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce an RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter HCRP encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding HCRP may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of HCRP activity, it may be useful to encode a chimeric HCRP protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the HCRP encoding sequence and the heterologous protein sequence, so that HCRP may be cleaved and purified away from the heterologous moiety.

In another embodiment, sequences encoding HCRP may be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232). Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of HCRP, or a fragment thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J.Y. et al. (1995) Science 269:202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer).

The newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography (e.g., Creighton, T. (1983) Proteins, Structures and Molecular Principles, WH Freeman and Co., New York, NY). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure; Creighton, supra). Additionally, the amino acid sequence of HCRP, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a biologically active HCRP, the nucleotide sequences encoding HCRP or functional equivalents, may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding HCRP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. Such techniques are described in Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, and Ausubel, F.M. et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY.

A variety of expression vector/host systems may be utilized to contain and express sequences encoding HCRP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems.

The invention is not limited by the host cell employed.

The "control elements" or "regulatory sequences" are those non-translated regions of the vector--enhancers, promoters, 5' and 3' untranslated regions--which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript® phagemid (Stratagene, LaJolla, CA) or pSport1™ plasmid (Gibco BRL) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO; and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding HCRP, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for HCRP. For example, when large quantities of HCRP are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional E. coli cloning and expression vectors such as Bluescript® (Stratagene), in which the sequence encoding HCRP may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of ß-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509); and the like. pGEX vectors (Promega, Madison, WI) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel et al. (supra) and Grant et al. (1987) Methods Enzymol. 153:516-544.

In cases where plant expression vectors are used, the expression of sequences encoding

HCRP may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs, S. or Murry, L.E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, NY; pp. 191-196.

An insect system may also be used to express HCRP. For example, in one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The sequences encoding HCRP may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of HCRP will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, S. frugiperda cells or Trichoplusia larvae in which HCRP may be expressed (Engelhard, E.K. et al. (1994) Proc. Nat. Acad. Sci. 91:3224-3227).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding HCRP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing HCRP in infected host cells (Logan, J. and Shenk, T. (1984) Proc. Natl. Acad. Sci. 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of 6 to 10M are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding HCRP. Such signals include the ATG initiation codon and adjacent

sequences. In cases where sequences encoding HCRP, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC; Bethesda, MD) and may be chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express HCRP may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler, M. et al. (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1980) Cell 22:817-23) genes which can be employed in tk^- or $aprt^-$ cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which

confers resistance to methotrexate (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-51). Recently, the use of visible markers has gained popularity with such markers as anthocyanins, ß glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. et al. (1995) Methods Mol. Biol. 55:121-131). .

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding HCRP is inserted within a marker gene sequence, transformed cells containing sequences encoding HCRP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding HCRP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which contain the nucleic acid sequence encoding HCRP and express HCRP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

The presence of polynucleotide sequences encoding HCRP can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or fragments or fragments of polynucleotides encoding HCRP. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding HCRP to detect transformants containing DNA or RNA encoding HCRP.

A variety of protocols for detecting and measuring the expression of HCRP, using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on HCRP is preferred, but a

competitive binding assay may be employed. These and other assays are described, among other places, in Hampton, R. et al. (1990; Serological Methods, a Laboratory Manual, APS Press, St Paul, MN) and Maddox, D.E. et al. (1983; J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding HCRP include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding HCRP, or any fragments thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, MI); Promega (Madison WI); and U.S. Biochemical Corp., Cleveland, OH). Suitable reporter molecules or labels, which may be used for ease of detection, include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding HCRP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode HCRP may be designed to contain signal sequences which direct secretion of HCRP through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join sequences encoding HCRP to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA) between the purification domain and HCRP may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing HCRP and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMAC (immobilized metal ion affinity

chromatography as described in Porath, J. et al. (1992, Prot. Exp. Purif. 3: 263-281) while the enterokinase cleavage site provides a means for purifying HCRP from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D.J. et al. (1993; DNA Cell Biol. 12:441-453).

In addition to recombinant production, fragments of HCRP may be produced by direct peptide synthesis using solid-phase techniques Merrifield J. (1963) J. Am. Chem. Soc. 85:2149-2154). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of HCRP may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

THERAPEUTICS

Chemical and structural homology exits among HCRP-1 and the cyclin-specific ubiquitin carrier protein E2-C from atlantic clam (GI 1493838) and the ubiquitin carrier protein E2-18KD from yeast (GI 171866). In addition, HCRP is expressed in cancerous tissues and tissues associated with inflammation and the immune response. Therefore, HCRP-1 appears to play a role in cancer and immune disorders. In particular, a decrease in the level or activity of HCRP-1 appears to be associated with the development of cancer or immune disorders.

Therefore, in one embodiment, HCRP-1 or a fragment or derivative thereof may be administered to a subject to prevent or treat cancer. Cancers may include, but are not limited to, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, and teratocarcinoma and particularly cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing HCRP-1, or a fragment or a derivative thereof, may also be administered to a subject to prevent or treat cancer, including those described above.

In still another embodiment, an agonist which modulates the activity of HCRP-1 may also be administered to a subject to prevent or treat cancer such as those described above.

In another embodiment, HCRP-1 or a fragment or derivative thereof may be administered to a subject to prevent or treat an immune disorder. Such disorders may include, but are not limited to, AIDS, Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitus, Crohn's disease, ulcerative colitis, atopic

dermatitis, dermatomyositis, diabetes mellitus, emphysema, atrophic gastritis, glomerulonephritis, gout, Graves' disease, hypereosinophilia. irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, and autoimmune thyroiditis; complications of cancer, hemodialysis, extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections and trauma.

In another embodiment, a vector capable of expressing HCRP-1, or a fragment or a derivative thereof, may also be administered to a subject to prevent or treat an immune disorder, including those described above.

In still another embodiment, an agonist which modulates the activity of HCRP-1 may also be administered to a subject to prevent or treat an immune disorder such as those described above.

Chemical and structural homology exits among HCRP-2 and cyclin B from mouse (GI 506137) and human (GI 105763). In addition, HCRP-2 is expressed in cancerous tissues and fetal tissues. Therefore, HCRP-2 appears to play a role in cancer and developmental disorders. In particular, an increase in the level or activity of HCRP-2 appears to be associated with cancer, while a decrease in the level or activity of HCRP-2 appears to be associated with developmental disorders.

Therefore, in another embodiment, HCRP-2 or a fragment or derivative thereof may be administered to a subject to prevent or treat a developmental disorder. The term "developmental disorder" refers to any disorder associated with development of an organ or organ system of a subject. i.e., adrenal gland, skeletal system, reproductive system, etc. Such disorders include, but are not limited to, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, epilepsy, gonadal dysgenesis, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spinal bifida, and congenital glaucoma, cataract, or sensorineural hearing loss.

In another embodiment, a vector capable of expressing HCRP-2, or a fragment or a derivative thereof, may also be administered to a subject to prevent or treat a developmental disorder, including those described above.

In still another embodiment, an agonist which modulates the activity of HCRP-2 may also be administered to a subject to prevent or treat a developmental disorder, including those

described above.

In another embodiment, antagonists which decrease the activity of HCRP-2 may be administered to a subject to prevent or treat cancer. Cancers may include, but are not limited to, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, and teratocarcinoma and particularly cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, antibodies which specifically bind HCRP-2 may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express HCRP-2.

In another embodiment, a vector expressing the complement of the polynucleotide encoding HCRP-2 may be administered to a subject to treat or prevent cancer, including those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

Antagonists or inhibitors of HCRP may be produced using methods which are generally known in the art. In particular, purified HCRP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind HCRP.

Antibodies to HCRP may be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies, (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others, may be immunized by injection with HCRP or any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as

lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to HCRP have an amino acid sequence consisting of at least five amino acids and more preferably at least 10 amino acids. It is also preferable that they are identical to a portion of the amino acid sequence of the natural protein, and they may contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of HCRP amino acids may be fused with those of another protein such as keyhole limpet hemocyanin and antibody produced against the chimeric molecule.

Monoclonal antibodies to HCRP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies", the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce HCRP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobin libraries (Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:11120-3).

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for HCRP may also be generated. For example, such fragments include, but are not limited to, the F(ab')2 fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively,

Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 254:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between HCRP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering HCRP epitopes is preferred, but a competitive binding assay may also be employed (Maddox, supra).

In another embodiment of the invention, the polynucleotides encoding HCRP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding HCRP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding HCRP. Thus, complementary molecules or fragments may be used to modulate HCRP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments, can be designed from various locations along the coding or control regions of sequences encoding HCRP.

Expression vectors derived from retro viruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids may be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Methods which are well known to those skilled in the art can be used to construct vectors which will express nucleic acid sequence which is complementary to the polynucleotides of the gene encoding HCRP. These techniques are described both in Sambrook et al. (supra) and in Ausubel et al. (supra).

Genes encoding HCRP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide or fragment thereof which encodes HCRP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector and even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5' or

regulatory regions of the gene encoding HCRP (signal sequence, promoters, enhancers, and introns). Oligonucleotides derived from the transcription initiation site, e.g., between positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) In: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY). The complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Examples which may be used include engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding HCRP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding HCRP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA constitutively or inducibly can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase

linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections or polycationic amino polymers (Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-66; incorporated herein by reference) may be achieved using methods which are well known in the art.

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of HCRP, antibodies to HCRP, mimetics, agonists, antagonists, or inhibitors of HCRP. The compositions may be administered alone or in combination with at least one other agent, such as stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for

delivery. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of HCRP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example HCRP or fragments thereof, antibodies of HCRP, agonists, antagonists or inhibitors of HCRP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and

animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind HCRP may be used for the diagnosis of conditions or diseases characterized by expression of HCRP, or in assays to monitor patients being treated with HCRP, agonists, antagonists or inhibitors. The antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for HCRP include methods which utilize the antibody and a label to detect HCRP in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules which are known in the art may be used, several of which are described above.

A variety of protocols including ELISA, RIA, and FACS for measuring HCRP are known in the art and provide a basis for diagnosing altered or abnormal levels of HCRP expression. Normal or standard values for HCRP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to HCRP

under conditions suitable for complex formation The amount of standard complex formation may be quantified by various methods, but preferably by photometric, means. Quantities of HCRP expressed in subject, control and disease, samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding HCRP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of HCRP may be correlated with disease. The diagnostic assay may be used to distinguish between absence, presence, and excess expression of HCRP, and to monitor regulation of HCRP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding HCRP or closely related molecules, may be used to identify nucleic acid sequences which encode HCRP. The specificity of the probe, whether it is made from a highly specific region, e.g., 10 unique nucleotides in the 5' regulatory region, or a less specific region, e.g., especially in the 3' coding region, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low) will determine whether the probe identifies only naturally occurring sequences encoding HCRP, alleles, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the HCRP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and derived from the nucleotide sequence of SEQ ID NO:2 or SEQ ID NO:4 or from genomic sequence including promoter, enhancer elements, and introns of the naturally occurring HCRP.

Means for producing specific hybridization probes for DNAs encoding HCRP include the cloning of nucleic acid sequences encoding HCRP or HCRP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, radionuclides such as 32P or 35S, or enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding HCRP may be used for the diagnosis of conditions, disorders, or diseases which are associated with expression of HCRP. Examples of such conditions or diseases include cancer, such as cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; immune disorders such as AIDS, Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitus, Crohn's disease, ulcerative colitis, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, atrophic gastritis, glomerulonephritis, gout, Graves' disease, hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, and autoimmune thyroiditis; complications of cancer, hemodialysis, extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections and trauma; and developmental disorders such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, epilepsy, gonadal dysgenesis, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spinal bifida, and congenital glaucoma, cataract, or sensorineural hearing loss. The polynucleotide sequences encoding HCRP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; or in dipstick, pin, ELISA assays or microarrays utilizing fluids or tissues from patient biopsies to detect altered HCRP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding HCRP may be useful in assays that detect activation or induction of various cancers, particularly those mentioned above. The nucleotide sequences encoding HCRP may be labeled by standard methods, and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the biopsied or extracted sample is significantly altered from that of a comparable control sample, the nucleotide sequences have hybridized with nucleotide sequences in the sample, and the presence of altered levels of nucleotide sequences encoding HCRP in the sample indicates the presence of the associated disease. Such assays may also be used to evaluate the efficacy of a particular therapeutic

treatment regimen in animal studies, in clinical trials, or in monitoring the treatment of an individual patient.

In order to provide a basis for the diagnosis of disease associated with expression of HCRP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, which encodes HCRP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with those from an experiment where a known amount of a substantially purified polynucleotide is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients who are symptomatic for disease. Deviation between standard and subject values is used to establish the presence of disease.

Once disease is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to evaluate whether the level of expression in the patient begins to approximate that which is observed in the normal patient. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding HCRP may involve the use of PCR. Such oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably consist of two nucleotide sequences, one with sense orientation (5'->3') and another with antisense (3'<-5'), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of HCRP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated (Melby, P.C. et al. (1993) J.

Immunol. Methods, 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides derived from any of the polynucleotide sequences described herein may be used as probes in microarrays. The microarrays can be used to monitor the expression level of large numbers of genes simultaneously (to produce a transcript image), and to identify genetic variants, mutations and polymorphisms. This information will be useful in determining gene function, understanding the genetic basis of disease, diagnosing disease, and in developing and monitoring the activity of therapeutic agents.

In one embodiment, the microarray is prepared and used according to the methods described in PCT application WO95/11995 (Chee et al.), Lockhart, D. J. et al. (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. et al. (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in their entirety by reference.

The microarray is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs fixed to a solid support. Microarrays may contain oligonucleotides which cover the known 5', or 3', sequence, or contain sequential oligonucleotides which cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray may be oligonucleotides that are specific to a gene or genes of interest in which at least a fragment of the sequence is known or that are specific to one or more unidentified cDNAs which are common to a particular cell type, developmental or disease state.

In order to produce oligonucleotides to a known sequence for a microarray, the gene of interest is examined using a computer algorithm which starts at the 5' or more preferably at the 3' end of the nucleotide sequence. The algorithm identifies oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, the oligomers may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT

application WO95/251116 (Baldeschweiler et al.) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array may be produced by hand or using available devises (slot blot or dot blot apparatus) materials and machines (including robotic instruments) and contain grids of 8 dots, 24 dots, 96 dots, 384 dots, 1536 dots or 6144 dots, or any other multiple which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using the microarrays, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray so that the probe sequences hybridize to complementary oligonucleotides of the microarray. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large scale correlation studies on the sequences, mutations, variants, or polymorphisms among samples.

In another embodiment of the invention, the nucleic acid sequences which encode HCRP may also be used to generate hybridization probes which are useful for mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome or to artificial chromosome constructions, such as human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions or single chromosome cDNA libraries as reviewed in Price, C.M. (1993) Blood Rev. 7:127-134, and Trask, B.J. (1991) Trends Genet. 7:149-154.

Fluorescent in situ hybridization (FISH as described in Verma et al. (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York, NY) may be

correlated with other physical chromosome mapping techniques and genetic map data. Examples of genetic map data can be found in various scientific journals or at Online Mendelian Inheritance in Man (OMIM). Correlation between the location of the gene encoding HCRP on a physical chromosomal map and a specific disease , or predisposition to a specific disease, may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier, or affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse. may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms, or parts thereof, by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti, R.A. et al. (1988) Nature 336:577-580), any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier, or affected individuals.

In another embodiment of the invention, HCRP, its catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between HCRP and the agent being tested, may be measured.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the protein of interest as described in published PCT application WO84/03564. In this method, as applied to HCRP large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with HCRP, or fragments thereof, and washed. Bound HCRP is then detected by methods well known in the art. Purified HCRP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding HCRP specifically compete with a test compound for binding HCRP. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with HCRP.

In additional embodiments, the nucleotide sequences which encode HCRP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

EXAMPLES

I cDNA Library Construction

COLNFET02

The COLNFET02 cDNA library was constructed from colon tissue obtained from a 20-week-old Caucasian female fetus. The pregnant mother was treated with erythromycin for seven days in the first trimester for bronchitis (specimen #RU95-10-0739; IIAM, Exton, PA).

The frozen tissue was homogenized and lysed using a Brinkmann Homogenizer Polytron PT-3000 (Brinkmann Instruments, Westbury, NJ) in guanidinium isothiocyanate solution. The lysate was centrifuged over a 5.7 M CsCl cushion using an Beckman SW28 rotor in a Beckman L8-70M Ultracentrifuge (Beckman Instruments) for 18 hours at 25,000 rpm at ambient temperature. The RNA was extracted with acid phenol pH 4.7, precipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, resuspended in RNAse-free water, and DNase treated at 37°C. RNA extraction and precipitation were repeated as before. The mRNA was then isolated using the Qiagen Oligotex kit (QIAGEN, Inc., Chatsworth, CA) and used to construct the cDNA library.

The mRNA was handled according to the recommended protocols in the SuperScript Plasmid System for cDNA Synthesis and Plasmid Cloning (Cat. #18248-013, Gibco/BRL, Gaithersburg, MD). cDNAs were fractionated on a Sepharose CL4B column (Cat. #275105-01, Pharmacia), and those cDNAs exceeding 400 bp were ligated into pSport I. The plasmid pSport I was subsequently transformed into DH5a™ competent cells (Cat. #18258-012, Gibco/BRL).

HMCINOT01

The human mast cell HMCINOT01 cDNA library was custom constructed by Stratagene

(Stratagene, La Jolla, CA 92037) using mRNA purified from cultured HMC-1 cells. The library was prepared by Stratagene essentially as described. The human mast cell (HMC-1) cDNA library was prepared by purifying poly(A+)RNA (mRNA) from human mast cells and then enzymatically synthesizing double stranded complementary DNA (cDNA) copies of the mRNA. Synthetic adaptor oligonucleotides were ligated onto the ends of the cDNA enabling its insertion into the lambda vector. The HMC-1 library was constructed using the Uni-ZAP™ vector system (Stratagene).

The HMC-1 cDNA library can be screened with either DNA probes or antibody probes and the pBluescript® phagemid (Stratagene) can be rapidly excised in vivo. The custom-constructed library phage particles were infected into E. coli host strain XL1-Blue® (Stratagene). Alternative unidirectional vectors might include, but are not limited to, pcDNAI (Invitrogen, San Diego, CA) and pSHlox-1 (Novagen, Madison WI).

II Isolation and Sequencing of cDNA Clones

The phagemid forms of individual cDNA clones were obtained by the in vivo excision process, in which the host bacterial strain was coinfected with both the lambda library phage and an f1 helper phage. Proteins derived from both the library-containing phage and the helper phage nicked the lambda DNA, initiated new DNA synthesis from defined sequences on the lambda target DNA and created a smaller, single stranded circular phagemid DNA molecule that included all DNA sequences of the pBluescript® plasmid and the cDNA insert. The phagemid DNA was secreted from the cells and purified, then used to re-infect fresh host cells, where the double stranded phagemid DNA was produced. Because the phagemid carries the gene for ß-lactamase, the newly-transformed bacteria are selected on medium containing ampicillin.

Phagemid DNA was purified using the Magic Minipreps™ DNA Purification System (catalogue #A7100. Promega Corp., Madison, WI 53711). The DNA was eluted from the purification resin already prepared for DNA sequencing and other analytical manipulations.

Phagemid DNA may also be purified using the QIAWELL-8 Plasmid or QIAGEN® DNA Purification System (QIAGEN Inc, Chatsworth, CA). The DNA was eluted from the purification resin and prepared for DNA sequencing and other analytical manipulations.

An alternative method for purifying phagemid DNA utilizes the Miniprep Kit available from Advanced Genetic Technologies Corp. (Gaithersburg, Maryland, Catalog No. 77468). This kit is in the 96-well format and provides enough reagents for 960 purifications. Each kit contains a recommended protocol, which is employed except for the following changes. First, the 96

wells are each filled with only 1 ml of sterile Terrific Broth (catalog #22711) with carbenicillin at 25 mg/L and glycerol at 0.4%. After the wells are inoculated, the bacteria are cultured for 24 hours and lysed with 60 μl of lysis buffer. A centrifugation step (2900 rpm for 5 minutes) is performed before the contents of the block is added to the primary filter plate. The optional step of adding isopropanol to TRIS buffer is not performed. After the last step in the protocol, samples are transferred to a Beckman 96-well block for storage.

Another altrenative method for purifying phagemid DNA uses the REAL Prep 96 Plasmid Kit (Catalog #26173, QIAGEN, Inc.). This kit enables the simultaneous purification of 96 samples in a 96-well block using multi-channel reagent dispensers. The recommended protocol is employed except for the following changes: 1) the bacteria are cultured in 1 ml of sterile Terrific Broth (Catalog #22711,Gibco/BRL) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) after inoculation, the cultures are incubated for 19 hours and at the end of incubation, the cells are lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet is resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples are transferred to a 96-well block for storage at 4° C.

The cDNAs were sequenced by the method of Sanger et al. (1975, J. Mol. Biol. 94:441f), using a Hamilton Micro Lab 2200 (Hamilton, Reno, NV) in combination with Peltier Thermal Cyclers (PTC200 from MJ Research, Watertown, MA) and Applied Biosystems 377 DNA Sequencing Systems; and the reading frame was determined.

III Homology Searching of cDNA Clones and Their Deduced Proteins

COLNFET02

The nucleotide sequences of the Sequence Listing or amino acid sequences deduced from them were used as query sequences against databases such as GenBank, SwissProt, BLOCKS, and Pima II. These databases which contain previously identified and annotated sequences were searched for regions of homology (similarity) using BLAST, which stands for Basic Local Alignment Search Tool (Altschul, S.F. (1993) J. Mol. Evol. 36:290-300; Altschul et al. (1990) J. Mol. Biol. 215:403-410).

BLAST produces alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST is especially useful in determining exact matches or in identifying homologs which may be of prokaryotic (bacterial) or eukaryotic (animal, fungal or plant) origin. Other algorithms such as the one described in Smith RF and TF Smith (1992; Protein Engineering 5:35-51), incorporated herein by reference, can be

used when dealing with primary sequence patterns and secondary structure gap penalties. As disclosed in this application, the sequences have lengths of at least 49 nucleotides, and no more than 12% uncalled bases (where N is recorded rather than A, C, G, or T).

The BLAST approach, as detailed in Karlin, S. and S.F. Atschul (1993; Proc. Nat. Acad. Sci. 90:5873-7) and incorporated herein by reference, searches for matches between a query sequence and a database sequence, to evaluate the statistical significance of any matches found, and to report only those matches which satisfy the user-selected threshold of significance. In this application, threshold was set at 10^{-25} for nucleotides and 10^{-14} for peptides.

Incyte nucleotide sequences were searched against the GenBank databases for primate (pri), rodent (rod), and mammalian sequences (mam), and deduced amino acid sequences from the same clones are searched against GenBank functional protein databases, mammalian (mamp), vertebrate (vrtp) and eukaryote (eukp), for homology. The relevant database for a particular match were reported as a GIxxx±p (where xxx is pri, rod, etc and if present, p = peptide).

HMCINOT01

Each cDNA was compared to sequences in GenBank using a search algorithm developed by Applied Biosystems and incorporated into the INHERIT™ 670 sequence analysis system. In this algorithm, Pattern Specification Language (TRW Inc, Los Angeles, CA) was used to determine regions of homology. The three parameters that determine how the sequence comparisons run were window size, window offset, and error tolerance. Using a combination of these three parameters, the DNA database was searched for sequences containing regions of homology to the query sequence, and the appropriate sequences were scored with an initial value. Subsequently, these homologous regions were examined using dot matrix homology plots to distinguish regions of homology from chance matches. Smith-Waterman alignments were used to display the results of the homology search.

Peptide and protein sequence homologies were ascertained using the INHERIT- 670 sequence analysis system using the methods similar to those used in DNA sequence homologies. Pattern Specification Language and parameter windows were used to search protein databases for sequences containing regions of homology which were scored with an initial value. Dot-matrix homology plots were examined to distinguish regions of significant homology from chance matches.

BLAST, which stands for Basic Local Alignment Search Tool (Altschul, S.F. (1993) J. Mol. Evol. 36:290-300; Altschul et al. (1990) J. Mol. Biol. 215:403-410), was used to search for local sequence alignments. BLAST produces alignments of both nucleotide and amino acid

sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST is especially useful in determining exact matches or in identifying homologs. BLAST is useful for matches which do not contain gaps. The fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP).

An HSP consists of two sequence fragments of arbitrary but equal lengths whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score set by the user. The BLAST approach is to look for HSPs between a query sequence and a database sequence, to evaluate the statistical significance of any matches found, and to report only those matches which satisfy the user-selected threshold of significance. The parameter E establishes the statistically significant threshold for reporting database sequence matches. E is interpreted as the upper bound of the expected frequency of chance occurrence of an HSP (or set of HSPs) within the context of the entire database search. Any database sequence whose match satisfies E is reported in the program output.

IV Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook et al., supra).

Analogous computer techniques using BLAST (Altschul, S.F. 1993 and 1990, supra) are used to search for identical or related molecules in nucleotide databases such as GenBank or the LIFESEQ™ database (Incyte Pharmaceuticals). This analysis is much faster than multiple, membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1-2% error; and at 70, the match will be exact. Homologous molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analysis are reported as a list of libraries in which the transcript encoding HCRP occurs. Abundance and percent abundance are also reported. Abundance

directly reflects the number of times a particular transcript is represented in a cDNA library, and percent abundance is abundance divided by the total number of sequences examined in the cDNA library.

V Extension of HCRP Encoding Polynucleotides

The nucleic acid sequence of the Incyte Clones 7755 and 1308478 were used to design oligonucleotide primers for extending a partial nucleotide sequence to full length. One primer was synthesized to initiate extension in the antisense direction, and the other was synthesized to extend sequence in the sense direction. Primers were used to facilitate the extension of the known sequence "outward" generating amplicons containing new, unknown nucleotide sequence for the region of interest. The initial primers were designed from the cDNA using OLIGO 4.06 (National Biosciences), or another appropriate program, to be about 22 to about 30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures of about 68°to about 72° C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries (Gibco/BRL) were used to extend the sequence If more than one extension is necessary or desired, additional sets of primers are designed to further extend the known region.

High fidelity amplification was obtained by following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix. Beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, PCR was performed using the Peltier Thermal Cycler (PTC200; M.J. Research, Watertown, MA) and the following parameters:

Step 1	94° C for 1 min (initial denaturation)
Step 2	65° C for 1 min
Step 3	68° C for 6 min
Step 4	94° C for 15 sec
Step 5	65° C for 1 min
Step 6	68° C for 7 min
Step 7	Repeat step 4-6 for 15 additional cycles
Step 8	94° C for 15 sec
Step 9	65° C for 1 min
Step 10	68° C for 7:15 min
Step 11	Repeat step 8-10 for 12 cycles
Step 12	72° C for 8 min
Step 13	4° C (and holding)

A 5-10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6-0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were excised from the gel, purified using QIAQuick™ (QIAGEN Inc., Chatsworth, CA), and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13 μl of ligation buffer, 1μl T4-DNA ligase (15 units) and 1μl T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2-3 hours or overnight at 16° C. Competent E. coli cells (in 40 μl of appropriate media) were transformed with 3 μl of ligation mixture and cultured in 80 μl of SOC medium (Sambrook et al., supra). After incubation for one hour at 37° C, the E. coli mixture was plated on Luria Bertani (LB)-agar (Sambrook et al., supra) containing 2x Carb. The following day, several colonies were randomly picked from each plate and cultured in 150 μl of liquid LB/2x Carb medium placed in an individual well of an appropriate, commercially-available, sterile 96-well microtiter plate. The following day, 5 μl of each overnight culture was transferred into a non-sterile 96-well plate and after dilution 1:10 with water, 5 μl of each sample was transferred into a PCR array.

For PCR amplification, 18 μl of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

Step 1	94° C for 60 sec
Step 2	94° C for 20 sec
Step 3	55° C for 30 sec
Step 4	72° C for 90 sec
Step 5	Repeat steps 2-4 for an additional 29 cycles
Step 6	72° C for 180 sec
Step 7	4° C (and holding)

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

In like manner, the nucleotide sequence of SEQ ID NO:2 or SEQ ID NO:4 is used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for 5' extension, and an appropriate genomic library.

VI Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:2 Oor SEQ ID NO:4 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base-pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 (National Biosciences), labeled by combining 50 pmol of each oligomer and 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham) and T4 polynucleotide kinase (DuPont NEN®, Boston, MA). The labeled oligonucleotides are substantially purified with Sephadex G-25 superfine resin column (Pharmacia & Upjohn). A aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases (Ase I, Bgl II, Eco RI, Pst I, Xba 1, or Pvu II; DuPont NEN®).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham, NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR™ film (Kodak, Rochester, NY) is exposed to the blots in a Phosphoimager cassette (Molecular Dynamics, Sunnyvale, CA) for several hours, hybridization patterns are compared visually.

VII Microarrays

To produce oligonucleotides for a microarray, the nucleotide sequence described herein is examined using a computer algorithm which starts at the 3' end of the nucleotide sequence. The algorithm identifies oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that would interfere with hybridization. The algorithm identifies 20 sequence-specific oligonucleotides of 20 nucleotides in length (20-mers). A matched set of oligonucleotides is created in which one nucleotide in the center of each sequence is altered. This process is repeated for each gene in the microarray, and double sets of twenty 20 mers are synthesized and arranged on the surface of the silicon chip using a light-directed chemical process (Chee, M. et al., PCT/WO95/11995, incorporated herein by reference).

In the alternative, a chemical coupling procedure and an ink jet device are used to synthesize oligomers on the surface of a substrate (Baldeschweiler, J.D. et al.,

PCT/WO95/25116, incorporated herein by reference). In another alternative, a "gridded" array analogous to a dot (or slot) blot is used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array may be produced by hand or using available materials and machines and contain grids of 8 dots, 24 dots, 96 dots, 384 dots, 1536 dots or 6144 dots. After hybridization, the microarray is washed to remove nonhybridized probes, and a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the micro-array.

VIII Complementary Polynucleotides

Sequence complementary to the HCRP-encoding sequence, or any part thereof, is used to decrease or inhibit expression of naturally occurring HCRP. Although use of oligonucleotides comprising from about 15 to about 30 base-pairs is described, essentially the same procedure is used with smaller or larger sequence fragments. Appropriate oligonucleotides are designed using Oligo 4.06 software and the coding sequence of HCRP, SEQ ID NO:1. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the HCRP-encoding transcript.

IX Expression of HCRP

Expression of HCRP is accomplished by subcloning the cDNAs into appropriate vectors and transforming the vectors into host cells. In this case, the cloning vector is also used to express HCRP in E. coli. Upstream of the cloning site, this vector contains a promoter for ß-galactosidase, followed by sequence containing the amino-terminal Met, and the subsequent seven residues of ß-galactosidase. Immediately following these eight residues is a bacteriophage promoter useful for transcription and a linker containing a number of unique restriction sites.

Induction of an isolated, transformed bacterial strain with IPTG using standard methods produces a fusion protein which consists of the first eight residues of ß-galactosidase, about 5 to 15 residues of linker, and the full length protein. The signal residues direct the secretion of HCRP into the bacterial growth media which can be used directly in the following assay for activity.

X Demonstration of HCRP Activity

HCRP-1

HCRP-1 activity is measured by the cyclin-ubiquitin ligation assay (Aristarkhov et al., supra). The reaction contains in a volume of 10 μl, 40 mM Tris.HCl (pH 7.6), 5 mM Mg Cl_2, 0.5 mM ATP, 10 mM phosphocreatine, 50 μg of creatine phosphokinase/ ml, 1 mg reduced carboxymethylated bovine serum albumin/ ml, 50 μM ubiquitin, 1 μM ubiquitin aldehyde, 1-2 pmol ^{125}I-labeled cyclin B, 1 pmol E1, 1 μM okadaic acid, 10 μg of protein of M-phase fraction 1A (containing active E3-C and essentially free of E2-C), and varying amounts of HCRP-1. The reaction is incubated at 18 °C for 60 minutes. Samples are then separated by electrophoresis on SDS/12% polyacrylamide gel. The amount of ^{125}I- cyclin-ubiquitin formed is quantified by PhosphorImager analysis. The amount of cyclin-ubiquitin formation is proportional to the amount of HCRP-1 in the reaction.

HCRP-2

HCRP-2 activity is measured in the same assay described above, in which variable amounts of ^{125}I-labeled HCRP-2 are substituted for the fixed amount of ^{125}I-labeled cyclin B used, and a fixed amount of purified E2-C is used in place of variable amounts of HCRP-1. Under these conditions, the amount of ubiquitin-HCRP-2 formation is proportional to the amount of HCRP-2 in the reaction.

XI Production of HCRP Specific Antibodies

HCRP that is substantially purified using PAGE electrophoresis (Sambrook, supra), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols. The amino acid sequence deduced from SEQ ID NO:2 is analyzed using DNASTAR software (DNASTAR Inc) to determine regions of high immunogenicity and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions, is described by Ausubel et al. (supra), and others.

Typically, the oligopeptides are 15 residues in length, synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry, and coupled to keyhole limpet hemocyanin (KLH, Sigma, St. Louis, MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS; Ausubel et al., supra). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. The resulting antisera are tested for antipeptide activity, for example, by binding the peptide to plastic, blocking with 1% BSA,

reacting with rabbit antisera, washing, and reacting with radio iodinated, goat anti-rabbit IgG.

XII Purification of Naturally Occurring HCRP Using Specific Antibodies

Naturally occurring or recombinant HCRP is substantially purified by immunoaffinity chromatography using antibodies specific for HCRP. An immunoaffinity column is constructed by covalently coupling HCRP antibody to an activated chromatographic resin, such as CNBr-activated Sepharose (Pharmacia & Upjohn). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing HCRP is passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of HCRP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/HCRP binding (eg, a buffer of pH 2-3 or a high concentration of a chaotrope, such as urea or thiocyanate ion), and HCRP is collected.

XIII Identification of Molecules Which Interact with HCRP

HCRP or biologically active fragments thereof are labeled with ^{125}I Bolton-Hunter reagent (Bolton et al. (1973) Biochem. J. 133: 529). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled HCRP, washed and any wells with labeled HCRP complex are assayed. Data obtained using different concentrations of HCRP are used to calculate values for the number, affinity, and association of HCRP with the candidate molecules.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: INCYTE PHARMACEUTICALS, INC.

(ii) TITLE OF THE INVENTION: CYCLIN RELATED PROTEINS

(iii) NUMBER OF SEQUENCES: 8

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Incyte Pharmaceuticals, Inc.
(B) STREET: 3174 Porter Drive
(C) CITY: Palo Alto
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 94304

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:
(A) PCT APPLICATION NUMBER: To Be Assigned
(B) FILING DATE: Filed Herewith

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/870,143
(B) FILING DATE: 05-JUN-1997

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Billings, Lucy J.
(B) REGISTRATION NUMBER: 36,749
(C) REFERENCE/DOCKET NUMBER: PF-0314 PCT

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 650-855-0555
(B) TELEFAX: 650-845-4166

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 179 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: COLNFET02
(B) CLONE: 1308478

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
Met Ala Ser Gln Asn Arg Asp Pro Ala Ala Thr Ser Val Ala Ala Ala
 1               5                  10                  15
Arg Lys Gly Ala Glu Pro Ser Gly Gly Ala Ala Arg Gly Pro Val Gly
            20                  25                  30
Lys Arg Leu Gln Gln Glu Leu Met Thr Leu Met Met Ser Gly Asp Lys
        35                  40                  45
Gly Ile Ser Ala Phe Pro Glu Ser Asp Asn Leu Phe Lys Trp Val Gly
    50                  55                  60
Thr Ile His Gly Ala Ala Gly Thr Val Tyr Glu Asp Leu Arg Tyr Lys
65                  70                  75                  80
Leu Ser Leu Glu Phe Pro Ser Gly Tyr Pro Tyr Asn Ala Pro Thr Val
```

```
                85                  90                  95
Lys Phe Leu Thr Pro Cys Tyr His Pro Asn Val Asp Thr Gln Gly Asn
            100                 105                 110
Ile Cys Leu Asp Ile Leu Lys Glu Lys Trp Ser Ala Leu Tyr Asp Val
        115                 120                 125
Arg Thr Ile Leu Leu Ser Ile Gln Ser Leu Leu Gly Glu Pro Asn Ile
    130                 135                 140
Asp Ser Pro Leu Asn Thr His Ala Ala Glu Leu Trp Lys Asn Pro Thr
145                 150                 155                 160
Ala Phe Lys Lys Tyr Leu Gln Glu Thr Tyr Ser Lys Gln Val Thr Ser
                165                 170                 175
Gln Glu Pro
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 837 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: COLNFET02
(B) CLONE: 1308478

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
CCCCAGAGGG TTACAATTCA AACGCGGGCG GGCGGGCCCG CAGTCCTGCA GTTGCAGTCG     60
TGTTCTCCGA GTTCCTGTCT CTCTGCCAAC GCCGCCCGGA TGGCTTCCCA AAACCGCGAC    120
CCAGCCGCCA CTAGCGTCGC CGCCGCCCGT AAAGGAGCTG AGCCGAGCGG GGGCGCCGCC    180
CGGGGTCCGG TGGGCAAAAG GCTACAGCAG GAGCTGATGA CCCTCATGAT GTCTGGCGAT    240
AAAGGGATTT CTGCCTTCCC TGAATCAGAC AACCTTTTCA AATGGGTAGG GACCATCCAT    300
GGAGCAGCTG GAACAGTATA TGAAGACCTG AGGTATAAGC TCTCGCTAGA GTTCCCCAGT    360
GGCTACCCTT ACAATGCGCC CACAGTGAAG TTCCTCACGC CCTGCTATCA CCCCAACGTG    420
GACACCCAGG GTAACATATG CCTGGACATC CTGAAGGAAA AGTGGTCTGC CCTGTATGAT    480
GTCAGGACCA TTCTGCTCTC CATCCAGAGC CTTCTAGGAG AACCCAACAT TGATAGTCCC    540
TTGAACACAC ATGCTGCCGA GCTCTGGAAA AACCCCACAG CTTTTAAGAA GTACCTGCAA    600
GAAACCTACT CAAAGCAGGT CACCAGCCAG GAGCCCTGAC CCAGGCTGCC CAGCCTGTCC    660
TTGTGTCGTC TTTTTAATTT TTCCTTAGAT GGTCTGTCCT TTTTGTGATT TCTGTATAGG    720
ACTCTTTATC TTGAGCTGTG GTATTTTTGT TTTGTTTTTG TCTTTTAAAT TAAGCCTCGG    780
TTGAGCCCTT GTATATTAAA TAAATGCATT TTTGTCCTTT TTTAGACAAA AAAAAAA       837
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 398 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: HMCINOT01
(B) CLONE: 7755

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Met Ala Leu Leu Arg Arg Pro Thr Val Ser Ser Asp Leu Glu Asn Ile
 1               5                  10                  15
Asp Thr Gly Val Asn Ser Lys Val Lys Ser His Val Thr Ile Arg Arg
            20                  25                  30
Thr Val Leu Glu Glu Ile Gly Asn Arg Val Thr Thr Arg Ala Ala Gln
        35                  40                  45
Val Ala Lys Lys Ala Gln Asn Thr Lys Val Pro Val Gln Pro Thr Lys
    50                  55                  60
Thr Thr Asn Val Asn Lys Gln Leu Lys Pro Thr Ala Ser Val Lys Pro
65                  70                  75                  80
```

```
Val Gln Met Glu Lys Leu Ala Pro Lys Gly Pro Ser Pro Thr Pro Glu
                85                  90                  95
Asp Val Ser Met Lys Glu Glu Asn Leu Cys Gln Ala Phe Ser Asp Ala
            100                 105                 110
Leu Leu Cys Lys Ile Glu Asp Ile Asp Asn Glu Asp Trp Glu Asn Pro
        115                 120                 125
Gln Leu Cys Ser Asp Tyr Val Lys Asp Ile Tyr Gln Tyr Leu Arg Gln
    130                 135                 140
Leu Glu Val Leu Gln Ser Ile Asn Pro His Phe Leu Asp Gly Arg Asp
145                 150                 155                 160
Ile Asn Gly Arg Met Arg Ala Ile Leu Val Asp Trp Leu Val Gln Val
                165                 170                 175
His Ser Lys Phe Arg Leu Leu Gln Glu Thr Leu Tyr Met Cys Val Gly
            180                 185                 190
Ile Met Asp Arg Phe Leu Gln Val Gln Pro Val Ser Arg Lys Lys Leu
        195                 200                 205
Gln Leu Val Gly Ile Thr Ala Leu Leu Leu Ala Ser Lys Tyr Glu Glu
    210                 215                 220
Met Phe Ser Pro Asn Ile Glu Asp Phe Val Tyr Ile Thr Asp Asn Ala
225                 230                 235                 240
Tyr Thr Ser Ser Gln Ile Arg Glu Met Glu Thr Leu Ile Leu Lys Glu
                245                 250                 255
Leu Lys Phe Glu Leu Gly Arg Pro Leu Pro Leu His Phe Leu Arg Arg
            260                 265                 270
Ala Ser Lys Ala Gly Glu Val Asp Val Glu Gln His Thr Leu Ala Lys
        275                 280                 285
Tyr Leu Met Glu Leu Thr Leu Ile Asp Tyr Asp Met Val His Tyr His
    290                 295                 300
Pro Ser Lys Val Ala Ala Ala Ala Ser Cys Leu Ser Gln Lys Val Leu
305                 310                 315                 320
Gly Gln Gly Lys Trp Asn Leu Lys Gln Gln Tyr Tyr Thr Gly Tyr Thr
                325                 330                 335
Glu Asn Glu Val Leu Glu Val Met Gln His Met Ala Lys Asn Val Val
            340                 345                 350
Lys Val Asn Glu Asn Leu Thr Lys Phe Ile Ala Ile Lys Asn Lys Tyr
        355                 360                 365
Ala Ser Ser Lys Leu Leu Lys Ile Ser Met Ile Pro Gln Leu Asn Ser
    370                 375                 380
Lys Ala Val Lys Asp Leu Ala Ser Pro Leu Ile Gly Arg Ser
385                 390                 395
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1981 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: HMCINOT01
(B) CLONE: 7755

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
AGCGTCGAAG ATCCCCAGCG CTGCGGGCTC GGAGAGCAGT CCTAACGGCG CCTCGTACGC     60
TAGTGTCCTC CCTTTTCAGT CCGCGTCCCT CCCTGGGCCG GGCTGGCACT CTTGCCTTCC    120
CCGTCCCTCA TGGCGCTGCT CCGACGCCCG ACGGTGTCCA GTGATTTGGA GAATATTGAC    180
ACAGGAGTTA ATTCTAAAGT TAAGAGTCAT GTGACTATTA GGCGAACTGT TTTAGAAGAA    240
ATTGGAAATA GAGTTACAAC CAGAGCAGCA CAAGTAGCTA AGAAAGCTCA GAACACCAAA    300
GTTCCAGTTC AACCCACCAA AACAACAAAT GTCAACAAAC AACTGAAACC TACTGCTTCT    360
GTCAAACCAG TACAGATGGA AAAGTTGGCT CCAAAGGGTC CTTCTCCCAC ACCTGAGGAT    420
GTCTCCATGA AGGAAGAGAA TCTCTGCCAA GCTTTTTCTG ATGCCTTGCT CTGCAAAATC    480
GAGGACATTG ATAACGAAGA TTGGGAGAAC CCTCAGCTCT GCAGTGACTA CGTTAAGGAT    540
ATCTATCAGT ATCTCAGGCA GCTGGAGGTT TTGCAGTCCA TAAACCCACA TTTCTTAGAT    600
GGAAGAGATA TAAATGGACG CATGCGTGCC ATCCTAGTGG ATTGGCTGGT ACAAGTCCAC    660
TCCAAGTTTA GGCTTCTGCA GGAGACTCTG TACATGTGCG TTGGCATTAT GGATCGATTT    720
```

```
TTACAGGTTC AGCCAGTTTC CCGGAAGAAG CTTCAATTAG TTGGGATTAC TGCTCTGCTC     780
TTGGCTTCCA AGTATGAGGA GATGTTTTCT CCAAATATTG AAGACTTTGT TTACATCACA     840
GACAATGCTT ATACCAGTTC CCAAATCCGA GAAATGGAAA CTCTAATTTT GAAAGAATTG     900
AAATTTGAGT TGGGTCGACC CTTGCCACTA CACTTCTTAA GGCGAGCATC AAAAGCCGGG     960
GAGGTTGATG TTGAACAGCA CACTTTAGCC AAGTATTTGA TGGAGCTGAC TCTCATCGAC    1020
TATGATATGG TGCATTATCA TCCTTCTAAG GTAGCAGCAG CTGCTTCCTG CTTGTCTCAG    1080
AAGGTTCTAG GACAAGGAAA ATGGAACTTA AAGCAGCAGT ATTACACAGG ATACACAGAG    1140
AATGAAGTAT TGGAAGTCAT GCAGCACATG GCCAAGAATG TGGTGAAAGT AAATGAAAAC    1200
TTAACTAAAT TCATCGCCAT CAAGAATAAG TATGCAAGCA GCAAACTCCT GAAGATCAGC    1260
ATGATCCCTC AGCTGAACTC AAAAGCCGTC AAAGACCTTG CCTCCCCACT GATAGGAAGG    1320
TCCTAGGCTG CCGTGGCCCC TGGGGATGTG TGCTTCATTG TGCCCTTTTT CTTATTGGTT    1380
TAGAACTCTT GATTTTGTAC ATAGTCCTCT GGTCTATCTC ATGAAACCTC TTCTCAGACC    1440
AGTTTTCTAA ACATATATTG AGGAAAAATA AAGCGATTGG TTTTTCTTAA AAAAAAAAAA    1500
AAAAAGGGGG CGGCCCCTCA AAAGGATCCC CCGGGGGGCC CATTTTTTCC CTTGCAGGCA    1560
ACTTCATCCC TCTCTCCAAA GGAAGAANNN NNNCTCTGGG GCTTCCCTGG TGGGGGAAGT    1620
TCTTGAAGGG TTTCCTGCTT CTCCTCTNTC CCACCCTGGT TCTGTTCCCT CGGCCATAAA    1680
TGTCCCAGGG CTTCTCTAGA TTTCTTGTGA ACCGAAAAAC CCCAGGGTTG GGTTTTCCCC    1740
TTTTCCCTTT CCTCACTCCC TCGGTTGGCC TCCCACACTT TCTTTTAATT TTTGGAAGGG    1800
TTTTGTTTTC TCTCCTCCCC TTCTCCCCTC AAACTCCTCC TCCCAAAGGG CCCATTTCAT    1860
TTGTAAATTC TCTGGCTCCC CCACCCCTCC TCTAATGTTT TTGGTGGATC AATTTTCTCC    1920
CTTCCCCTCT TGTTTAATTG TGCTTTACCA ACCCCCCCCC CAGCNCAGCG TCTCCTTACT    1980
C                                                                    1981
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 177 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: GenBank
(B) CLONE: 1493838

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
Met Ser Gly Gln Asn Ile Asp Pro Ala Ala Asn Gln Val Arg Gln Lys
 1               5                  10                  15
Glu Arg Pro Arg Asp Met Thr Thr Ser Lys Glu Arg His Ser Val Ser
            20                  25                  30
Lys Arg Leu Gln Gln Glu Leu Arg Thr Leu Leu Met Ser Gly Asp Pro
        35                  40                  45
Gly Ile Thr Ala Phe Pro Asp Gly Asp Asn Leu Phe Lys Trp Val Ala
    50                  55                  60
Thr Leu Asp Gly Pro Lys Asp Thr Val Tyr Glu Ser Leu Lys Tyr Lys
65                  70                  75                  80
Leu Thr Leu Glu Phe Pro Ser Asp Tyr Pro Tyr Lys Pro Pro Val Val
                85                  90                  95
Lys Phe Thr Thr Pro Cys Trp His Pro Asn Val Asp Gln Ser Gly Asn
            100                 105                 110
Ile Cys Leu Asp Ile Leu Lys Glu Asn Trp Thr Ala Ser Tyr Asp Val
        115                 120                 125
Arg Thr Ile Leu Leu Ser Leu Gln Ser Leu Leu Gly Glu Pro Asn Asn
    130                 135                 140
Ala Ser Pro Leu Asn Ala Gln Ala Ala Asp Met Trp Ser Asn Gln Thr
145                 150                 155                 160
Glu Tyr Lys Lys Val Leu His Glu Lys Tyr Lys Thr Ala Gln Ser Asp
                165                 170                 175
Lys
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 156 amino acids
(B) TYPE: amino acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: GenBank
(B) CLONE: 1717866

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Ala Val Glu Glu Gly Gly Cys Val Thr Lys Arg Leu Gln Asn Glu
 1               5                  10                  15
Leu Leu Gln Leu Leu Ser Ser Thr Thr Glu Ser Ile Ser Ala Phe Pro
            20                  25                  30
Val Asp Asp Asn Asp Leu Thr Tyr Trp Val Gly Tyr Ile Thr Gly Pro
        35                  40                  45
Lys Asp Thr Pro Tyr Ser Gly Leu Lys Phe Lys Val Ser Leu Lys Phe
    50                  55                  60
Pro Gln Asn Tyr Pro Phe His Pro Pro Met Ile Lys Phe Leu Ser Pro
65                  70                  75                  80
Met Trp His Pro Asn Val Asp Lys Ser Gly Asn Ile Cys Leu Asp Ile
                85                  90                  95
Leu Lys Glu Lys Trp Ser Ala Val Tyr Asn Val Glu Thr Ile Leu Leu
            100                 105                 110
Ser Leu Gln Ser Leu Leu Gly Glu Pro Asn Asn Arg Ser Pro Leu Asn
        115                 120                 125
Ala Val Ala Ala Glu Leu Trp Asp Ala Asp Met Glu Glu Tyr Arg Lys
    130                 135                 140
Lys Val Leu Ala Cys Tyr Glu Glu Ile Asp Asp Tyr
145                 150                 155
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 398 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: GenBank
(B) CLONE: 50613

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
Met Ala Leu Leu Arg Arg Pro Thr Val Ser Ser Asp Leu Lys Asn Ile
 1               5                  10                  15
Asp Thr Glu Val Ile Pro Lys Ala Lys Ser His Val Thr Ile Arg Arg
            20                  25                  30
Ala Val Leu Glu Glu Ile Gly Asn Lys Val Arg Asn Arg Thr Thr Gln
        35                  40                  45
Val Ala Lys Lys Pro Gln Asn Thr Lys Val Pro Ala Leu Ser Thr Lys
    50                  55                  60
Val Thr Asn Val Asn Lys Gln Pro Lys Pro Thr Ala Ser Val Lys Pro
65                  70                  75                  80
Val Gln Met Glu Ala Leu Ala Pro Lys Asp Arg Pro Pro Ala Pro Glu
                85                  90                  95
Asp Val Ser Met Lys Glu Glu Ser Leu Cys Gln Ala Phe Ser Asp Ala
            100                 105                 110
Leu Leu Cys Lys Ile Glu Asp Ile Asp Asn Glu Asp Arg Glu Asn Pro
        115                 120                 125
Gln Leu Cys Ser Asp Tyr Val Lys Asp Ile Tyr Gln Tyr Leu Arg Gln
    130                 135                 140
Leu Glu Val Leu Gln Ser Ile Asn Pro His Phe Leu Asp Gly Arg Asp
145                 150                 155                 160
```

```
Ile Asn Gly Arg Met Arg Ala Ile Leu Val Asp Trp Leu Val Gln Val
                165                 170                 175
His Ser Lys Phe Arg Leu Leu Gln Glu Thr Leu Tyr Met Cys Ile Gly
            180                 185                 190
Ile Met Asp Arg Phe Leu Gln Ala Gln Leu Val Cys Arg Lys Lys Leu
        195                 200                 205
Gln Val Val Gly Ile Thr Ala Leu Leu Leu Ala Ser Lys Tyr Glu Glu
    210                 215                 220
Met Phe Ser Pro Asn Ile Glu Asp Phe Val Tyr Ile Thr Asp Asn Ala
225                 230                 235                 240
Tyr Thr Ser Ser Gln Ile Arg Glu Met Glu Thr Leu Ile Leu Lys Glu
                245                 250                 255
Leu Lys Phe Glu Leu Gly Arg Pro Leu Pro Leu His Phe Leu Arg Arg
            260                 265                 270
Ala Ser Lys Ala Gly Glu Val Asp Val Glu Gln His Thr Leu Ala Lys
        275                 280                 285
Tyr Leu Met Glu Leu Thr Leu Val Asp Tyr Asp Met Val His Tyr His
    290                 295                 300
Pro Ser Gln Val Ala Ala Ala Ala Ser Cys Leu Ser Gln Lys Val Leu
305                 310                 315                 320
Gly Gln Gly Lys Trp Asn Leu Lys Gln Gln Tyr Tyr Thr Gly Tyr Met
                325                 330                 335
Glu Ser Glu Val Leu Glu Val Met Gln His Met Ala Lys Asn Val Val
            340                 345                 350
Lys Val Asn Asp Asn Arg Thr Lys Phe Ile Ala Val Lys Asn Lys Tyr
        355                 360                 365

Ala Ser Ser Arg Leu Leu Lys Ile Lys His Asp Pro Gln Leu Asn Ser
    370                 375                 380
Lys Ile Ile Lys Asp Leu Val Ser Pro Leu Leu Gly Ser Pro
385                 390                 395
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 433 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: GenBank
(B) CLONE: 105673

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Met Ala Leu Arg Val Thr Arg Asn Ser Lys Ile Asn Ala Glu Asn Lys
 1               5                  10                  15
Ala Lys Ile Asn Met Ala Gly Ala Lys Arg Val Pro Thr Ala Pro Ala
            20                  25                  30
Ala Thr Ser Lys Pro Gly Leu Arg Pro Arg Thr Ala Leu Gly Asp Ile
        35                  40                  45
Gly Asn Lys Val Ser Glu Gln Leu Gln Ala Lys Met Pro Met Lys Lys
    50                  55                  60
Glu Ala Lys Pro Ser Ala Thr Gly Lys Val Ile Asp Lys Lys Leu Pro
65                  70                  75                  80
Lys Pro Leu Glu Lys Val Pro Met Leu Val Pro Val Pro Val Ser Glu
                85                  90                  95
Pro Val Pro Glu Pro Glu Pro Glu Pro Glu Pro Glu Pro Val Lys Glu
            100                 105                 110
Glu Lys Leu Ser Pro Glu Pro Ile Leu Val Asp Thr Ala Ser Pro Ser
        115                 120                 125
Pro Met Glu Thr Ser Gly Cys Ala Pro Ala Glu Glu Asp Leu Cys Gln
    130                 135                 140
Ala Phe Ser Asp Val Ile Leu Ala Val Asn Asp Val Asp Ala Glu Asp
145                 150                 155                 160
Gly Ala Asp Pro Asn Leu Cys Ser Glu Tyr Val Lys Asp Ile Tyr Ala
```

```
                 165                 170                 175
Tyr Leu Arg Gln Leu Glu Glu Glu Gln Ala Val Arg Pro Lys Tyr Leu
             180                 185                 190
Leu Gly Arg Glu Val Thr Gly Asn Met Arg Ala Ile Leu Ile Asp Trp
         195                 200                 205
Leu Val Gln Val Gln Met Lys Phe Arg Leu Leu Gln Glu Thr Met Tyr
     210                 215                 220
Met Thr Val Ser Ile Ile Asp Arg Phe Met Gln Asn Asn Cys Val Pro
225                 230                 235                 240
Lys Lys Met Leu Gln Leu Val Gly Val Thr Ala Met Phe Ile Ala Ser
                245                 250                 255
Lys Tyr Glu Glu Met Tyr Pro Pro Glu Ile Gly Asp Phe Ala Phe Val
            260                 265                 270
Thr Asp Asn Thr Tyr Thr Lys His Gln Ile Arg Gln Met Glu Met Lys
        275                 280                 285
Ile Leu Arg Ala Leu Asn Phe Gly Leu Gly Arg Pro Leu Pro Leu His
    290                 295                 300
Phe Leu Arg Arg Ala Ser Lys Ile Gly Glu Val Asp Val Glu Gln His
305                 310                 315                 320
Thr Leu Ala Lys Tyr Leu Met Glu Leu Thr Met Leu Asp Tyr Asp Met
                325                 330                 335
Val His Phe Pro Pro Ser Gln Ile Ala Ala Gly Ala Phe Cys Leu Ala
            340                 345                 350
Leu Lys Ile Leu Asp Asn Gly Glu Trp Thr Pro Thr Leu Gln His Tyr
        355                 360                 365
Leu Ser Tyr Thr Glu Glu Ser Leu Leu Pro Val Met Gln His Leu Ala
    370                 375                 380
Lys Asn Val Val Met Val Asn Gln Gly Leu Thr Lys His Met Thr Val
385                 390                 395                 400
Lys Asn Lys Tyr Ala Thr Ser Lys His Ala Lys Ile Ser Thr Leu Pro
                405                 410                 415
Gln Leu Asn Ser Ala Leu Val Gln Asp Leu Ala Lys Ala Val Ala Lys
            420                 425                 430
Val
```

What is claimed is:

1. A substantially purified cyclin related protein comprising the amino acid sequence of SEQ ID NO:1 or fragments thereof.

2. An isolated and purified polynucleotide sequence encoding the cyclin related protein of claim 1 or fragments or variants of said polynucleotide sequence.

3. A composition comprising the polynucleotide sequence of claim 2.

4. A polynucleotide sequence which hybridizes under stringent conditions to the polynucleotide sequence of claim 2.

5. A polynucleotide sequence which is complementary to the polynucleotide sequence of claim 2 or fragments or variants thereof.

6 An isolated and purified polynucleotide sequence comprising SEQ ID NO:2 or fragments or variants thereof.

7. A composition comprising the polynucleotide sequence of claim 6.

8. A polynucleotide sequence which is complementary to the polynucleotide sequence of claim 6.

9. An expression vector containing at least a fragment of the polynucleotide sequence of claim 2.

10. A host cell containing the vector of claim 9.

11. A method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:1, or a fragment thereof, the method comprising the steps of:

a) culturing the host cell of claim 10 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

12. A pharmaceutical composition comprising a substantially purified cyclin related protein having the amino acid sequence of SEQ ID NO:1 in conjunction with a suitable pharmaceutical carrier.

13. A purified antibody which specifically binds to the polypeptide of claim 1.

14. A purified agonist which modulates the activity of the polypeptide of claim 1.

15. A method for treating cancer comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 12.

16. A method for treating an immune disorder comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 12.

17. A method for detecting a polynucleotide which encodes a cyclin related protein in a biological sample comprising the steps of:

a) hybridizing the polynucleotide of claim 8 to nucleic acid material of a biological sample, thereby forming a hybridization complex; and

b) detecting said hybridization complex, wherein the presence of said complex correlates with the presence of a polynucleotide encoding cyclin related protein in said biological sample.

18. The method of claim 17 wherein the nucleic acid material is amplified by the polymerase chain reaction.

19. A substantially purified cyclin related protein comprising the amino acid sequence of SEQ ID NO:3 or fragments thereof.

20. An isolated and purified polynucleotide sequence encoding the cyclin related protein of claim 19 or fragments or variants of said polynucleotide sequence.

21. A composition comprising the polynucleotide sequence of claim 20.

22. A polynucleotide sequence which hybridizes under stringent conditions to the polynucleotide sequence of claim 20.

23. A polynucleotide sequence which is complementary to the polynucleotide sequence of claim 20 or fragments or variants thereof.

24. An isolated and purified polynucleotide sequence comprising SEQ ID NO:4 or fragments or variants thereof.

25. A composition comprising the polynucleotide sequence of claim 24.

26. A polynucleotide sequence which is complementary to the polynucleotide sequence of claim 24.

27. An expression vector containing at least a fragment of the polynucleotide sequence of claim 20.

28. A host cell containing the vector of claim 27.

29. A method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:3, or a fragment thereof, the method comprising the steps of:

a) culturing the host cell of claim 28 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

30. A pharmaceutical composition comprising a substantially purified cyclin related protein having the amino acid sequence of SEQ ID NO:3 in conjunction with a suitable pharmaceutical carrier.

31. A purified antibody which specifically binds to the polypeptide of claim 19.

32. A purified agonist which modulates the activity of the polypeptide of claim 19.

33. A purified antagonist which decreases the effect of the polypeptide of claim 19.

34. A method for treating a developmental disorder comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 30.

35. A method for treating an immune disorder comprising administering to a subject in need of such treatment an effective amount of the purified antagonist of claim 33.

36. A method for detecting a polynucleotide which encodes a cyclin related protein in a biological sample comprising the steps of:

a) hybridizing the polynucleotide of claim 26 to nucleic acid material of a biological sample, thereby forming a hybridization complex; and

b) detecting said hybridization complex, wherein the presence of said complex correlates with the presence of a polynucleotide encoding cyclin related protein in said biological sample.

37. The method of claim 36 wherein the nucleic acid material is amplified by the polymerase chain reaction.

```
          9          18          27          36          45          54
CCC CAG AGG GTT ACA ATT CAA ACG CGG GCG GGC GGG CCC GCA GTC CTG CAG TTG

          63          72          81          90          99         108
CAG TCG TGT TCT CCG AGT TCC TGT CTC TCT GCC AAC GCC GCC CGG ATG GCT TCC
                                                          M   A   S

         117         126         135         144         153         162
CAA AAC CGC GAC CCA GCC GCC ACT AGC GTC GCC GCC GCC CGT AAA GGA GCT GAG
Q   N   R   D   P   A   A   T   S   V   A   A   A   R   K   G   A   E

         171         180         189         198         207         216
CCG AGC GGG GGC GCC GCC CGG GGT CCG GTG GGC AAA AGG CTA CAG CAG GAG CTG
P   S   G   G   A   A   R   G   P   V   G   K   R   L   Q   Q   E   L

         225         234         243         252         261         270
ATG ACC CTC ATG ATG TCT GGC GAT AAA GGG ATT TCT GCC TTC CCT GAA TCA GAC
M   T   L   M   M   S   G   D   K   G   I   S   A   F   P   E   S   D

         279         288         297         306         315         324
AAC CTT TTC AAA TGG GTA GGG ACC ATC CAT GGA GCA GCT GGA ACA GTA TAT GAA
N   L   F   K   W   V   G   T   I   H   G   A   A   G   T   V   Y   E
```

FIGURE 1A

```
          333        342        351        360        369        378
GAC CTG AGG TAT AAG CTC TCG CTA GAG TTC CCC AGT GGC TAC CCT TAC AAT GCG
D   L   R   Y   K   L   S   L   E   F   P   S   G   Y   P   Y   N   A

          387        396        405        414        423        432
CCC ACA GTG AAG TTC CTC ACG CCC TGC TAT CAC CCC AAC GTG GAC ACC CAG GGT
P   T   V   K   F   L   T   P   C   Y   H   P   N   V   D   T   Q   G

          441        450        459        468        477        486
AAC ATA TGC CTG GAC ATC CTG AAG GAA AAG TGG TCT GCC CTG TAT GAT GTC AGG
N   I   C   L   D   I   L   K   E   K   W   S   A   L   Y   D   V   R

          495        504        513        522        531        540
ACC ATT CTG CTC TCC ATC CAG AGC CTT CTA GGA GAA CCC AAC ATT GAT AGT CCC
T   I   L   L   S   I   Q   S   L   L   G   E   P   N   I   D   S   P

          549        558        567        576        585        594
TTG AAC ACA CAT GCT GCC GAG CTC TGG AAA AAC CCC ACA GCT TTT AAG AAG TAC
L   N   T   H   A   A   E   L   W   K   N   P   T   A   F   K   K   Y

          603        612        621        630        639        648
CTG CAA GAA ACC TAC TCA AAG CAG GTC ACC AGC CAG GAG CCC TGA CCC AGG CTG
L   Q   E   T   Y   S   K   Q   V   T   S   Q   E   P
```

FIGURE 1B

```
              657             666             675             684             693             702
CCC AGC CTG TCC TTG TGT CGT CTT TTT AAT TTT TCC TTA GAT GGT CTG TCC TTT

              711             720             729             738             747             756
TTG TGA TTT CTG TAT AGG ACT CTT TAT CTT GAG CTG TGG TAT TTT TGT TTT GTT

              765             774             783             792             801             810
TTT GTC TTT TAA ATT AAG CCT CGG TTG AGC CCT TGT ATA TTA AAT AAA TGC ATT

              819             828             837
TTT GTC CTT TTT TAG ACA AAA AAA AAA
```

FIGURE 1C

```
          9          18          27          36          45          54
AGC GTC GAA GAT CCC CAG CGC TGC GGG CTC GGA GAG CAG TCC TAA CGG CGC CTC

         63          72          81          90          99         108
GTA CGC TAG TGT CCT CCC TTT TCA GTC CGC GTC CCT CCC TGG GCC GGG CTG GCA

        117         126         135         144         153         162
CTC TTG CCT TCC CCG TCC CTC ATG GCG CTG CTC CGA CGC CCG ACG GTG TCC AGT
                        M   A   L   L   R   R   P   T   V   S   S

        171         180         189         198         207         216
GAT TTG GAG AAT ATT GAC ACA GGA GTT AAT TCT AAA GTT AAG AGT CAT GTG ACT
D   L   E   N   I   D   T   G   V   N   S   K   V   K   S   H   V   T

        225         234         243         252         261         270
ATT AGG CGA ACT GTT TTA GAA GAA ATT GGA AAT AGA GTT ACA ACC AGA GCA GCA
I   R   R   T   V   L   E   E   I   G   N   R   V   T   T   R   A   A

        279         288         297         306         315         324
CAA GTA GCT AAG AAA GCT CAG AAC ACC AAA GTT CCA GTT CAA CCC ACC AAA ACA
Q   V   A   K   K   A   Q   N   T   K   V   P   V   Q   P   T   K   T
```

FIGURE 2A

```
          333        342        351        360        369        378
ACA AAT GTC AAC AAA CAA CTG AAA CCT ACT GCT TCT GTC AAA CCA GTA CAG ATG
T   N   V   N   K   Q   L   K   P   T   A   S   V   K   P   V   Q   M

          387        396        405        414        423        432
GAA AAG TTG GCT CCA AAG GGT CCT TCT CCC ACA CCT GAG GAT GTC TCC ATG AAG
E   K   L   A   P   K   G   P   S   P   T   P   E   D   V   S   M   K

          441        450        459        468        477        486
GAA GAG AAT CTC TGC CAA GCT TTT TCT GAT GCC TTG CTC TGC AAA ATC GAG GAC
E   E   N   L   C   Q   A   F   S   D   A   L   L   C   K   I   E   D

          495        504        513        522        531        540
ATT GAT AAC GAA GAT TGG GAG AAC CCT CAG CTC TGC AGT GAC TAC GTT AAG GAT
I   D   N   E   D   W   E   N   P   Q   L   C   S   D   Y   V   K   D

          549        558        567        576        585        594
ATC TAT CAG TAT CTC AGG CAG CTG GAG GTT TTG CAG TCC ATA AAC CCA CAT TTC
I   Y   Q   Y   L   R   Q   L   E   V   L   Q   S   I   N   P   H   F

          603        612        621        630        639        648
TTA GAT GGA AGA GAT ATA AAT GGA CGC ATG CGT GCC ATC CTA GTG GAT TGG CTG
L   D   G   R   D   I   N   G   R   M   R   A   I   L   V   D   W   L
```

FIGURE 2B

```
          657        666        675        684        693        702
GTA CAA GTC CAC TCC AAG TTT AGG CTT CTG CAG GAG ACT CTG TAC ATG TGC GTT
V   Q   V   H   S   K   F   R   L   L   Q   E   T   L   Y   M   C   V

          711        720        729        738        747        756
GGC ATT ATG GAT CGA TTT TTA CAG GTT CAG CCA GTT TCC CGG AAG AAG CTT CAA
G   I   M   D   R   F   L   Q   V   Q   P   V   S   R   K   K   L   Q

          765        774        783        792        801        810
TTA GTT GGG ATT ACT GCT CTG CTC TTG GCT TCC AAG TAT GAG GAG ATG TTT TCT
L   V   G   I   T   A   L   L   L   A   S   K   Y   E   E   M   F   S

          819        828        837        846        855        864
CCA AAT ATT GAA GAC TTT GTT TAC ATC ACA GAC AAT GCT TAT ACC AGT TCC CAA
P   N   I   E   D   F   V   Y   I   T   D   N   A   Y   T   S   S   Q

          873        882        891        900        909        918
ATC CGA GAA ATG GAA ACT CTA ATT TTG AAA GAA TTG AAA TTT GAG TTG GGT CGA
I   R   E   M   E   T   L   I   L   K   E   L   K   F   E   L   G   R

          927        936        945        954        963        972
CCC TTG CCA CTA CAC TTC TTA AGG CGA GCA TCA AAA GCC GGG GAG GTT GAT GTT
P   L   P   L   H   F   L   R   R   A   S   K   A   G   E   V   D   V
```

FIGURE 2C

```
         981        990        999       1008       1017       1026
GAA CAG CAC ACT TTA GCC AAG TAT TTG ATG GAG CTG ACT CTC ATC GAC TAT GAT
E   Q   H   T   L   A   K   Y   L   M   E   L   T   L   I   D   Y   D

        1035       1044       1053       1062       1071       1080
ATG GTG CAT TAT CAT CCT TCT AAG GTA GCA GCA GCT GCT TCC TGC TTG TCT CAG
M   V   H   Y   H   P   S   K   V   A   A   A   A   S   C   L   S   Q

        1089       1098       1107       1116       1125       1134
AAG GTT CTA GGA CAA GGA AAA TGG AAC TTA AAG CAG CAG TAT TAC ACA GGA TAC
K   V   L   G   Q   G   K   W   N   L   K   Q   Q   Y   Y   T   G   Y

        1143       1152       1161       1170       1179       1188
ACA GAG AAT GAA GTA TTG GAA GTC ATG CAG CAC ATG GCC AAG AAT GTG GTG AAA
T   E   N   E   V   L   E   V   M   Q   H   M   A   K   N   V   V   K

        1197       1206       1215       1224       1233       1242
GTA AAT GAA AAC TTA ACT AAA TTC ATC GCC ATC AAG AAT AAG TAT GCA AGC AGC
V   N   E   N   L   T   K   F   I   A   I   K   N   K   Y   A   S   S

        1251       1260       1269       1278       1287       1296
AAA CTC CTG AAG ATC AGC ATG ATC CCT CAG CTG AAC TCA AAA GCC GTC AAA GAC
K   L   L   K   I   S   M   I   P   Q   L   N   S   K   A   V   K   D
```

FIGURE 2D

```
      1305        1314        1323        1332        1341        1350
CTT GCC TCC CCA CTG ATA GGA AGG TCC TAG GCT GCC GTG GCC CCT GGG GAT GTG
L   A   S   P   L   I   G   R   S

      1359        1368        1377        1386        1395        1404
TGC TTC ATT GTG CCC TTT TTC TTA TTG GTT TAG AAC TCT TGA TTT TGT ACA TAG

      1413        1422        1431        1440        1449        1458
TCC TCT GGT CTA TCT CAT GAA ACC TCT TCT CAG ACC AGT TTT CTA AAC ATA TAT

      1467        1476        1485        1494        1503        1512
TGA GGA AAA ATA AAG CGA TTG GTT TTT CTT AAA AAA AAA AAA AAA AAG GGG GCG

      1521        1530        1539        1548        1557        1566
GCC CCT CAA AAG GAT CCC CCG GGG GGC CCA TTT TTT CCC TTG CAG GCA ACT TCA

      1575        1584        1593        1602        1611        1620
TCC CTC TCT CCA AAG GAA GAA NNN NNN CTC TGG GGC TTC CCT GGT GGG GGA AGT
```

FIGURE 2E

```
      1629        1638        1647        1656        1665        1674
TCT TGA AGG GTT TCC TGC TTC TCC TCT NTC CCA CCC TGG TTC TGT TCC CTC GGC

      1683        1692        1701        1710        1719        1728
CAT AAA TGT CCC AGG GCT TCT CTA GAT TTC TTG TGA ACC GAA AAA CCC CAG GGT

      1737        1746        1755        1764        1773        1782
TGG GTT TTC CCC TTT TCC CTT TCC TCA CTC CCT CGG TTG GCC TCC CAC ACT TTC

      1791        1800        1809        1818        1827        1836
TTT TAA TTT TTG GAA GGG TTT TGT TTT CTC TCC TCC CCT TCT CCC CTC AAA CTC

      1845        1854        1863        1872        1881        1890
CTC CTC CCA AAG GGC CCA TTT CAT TTG TAA ATT CTC TGG CTC CCC CAC CCC TCC

      1899        1908        1917        1926        1935        1944
TCT AAT GTT TTT GGT GGA TCA ATT TTC TCC CTT CCC CTC TTG TTT AAT TGT GCT

      1953        1962        1971        1980
TTA CCA ACC CCC CCC CCA GCN CAG CGT CTC CTT ACT C
```

FIGURE 2F

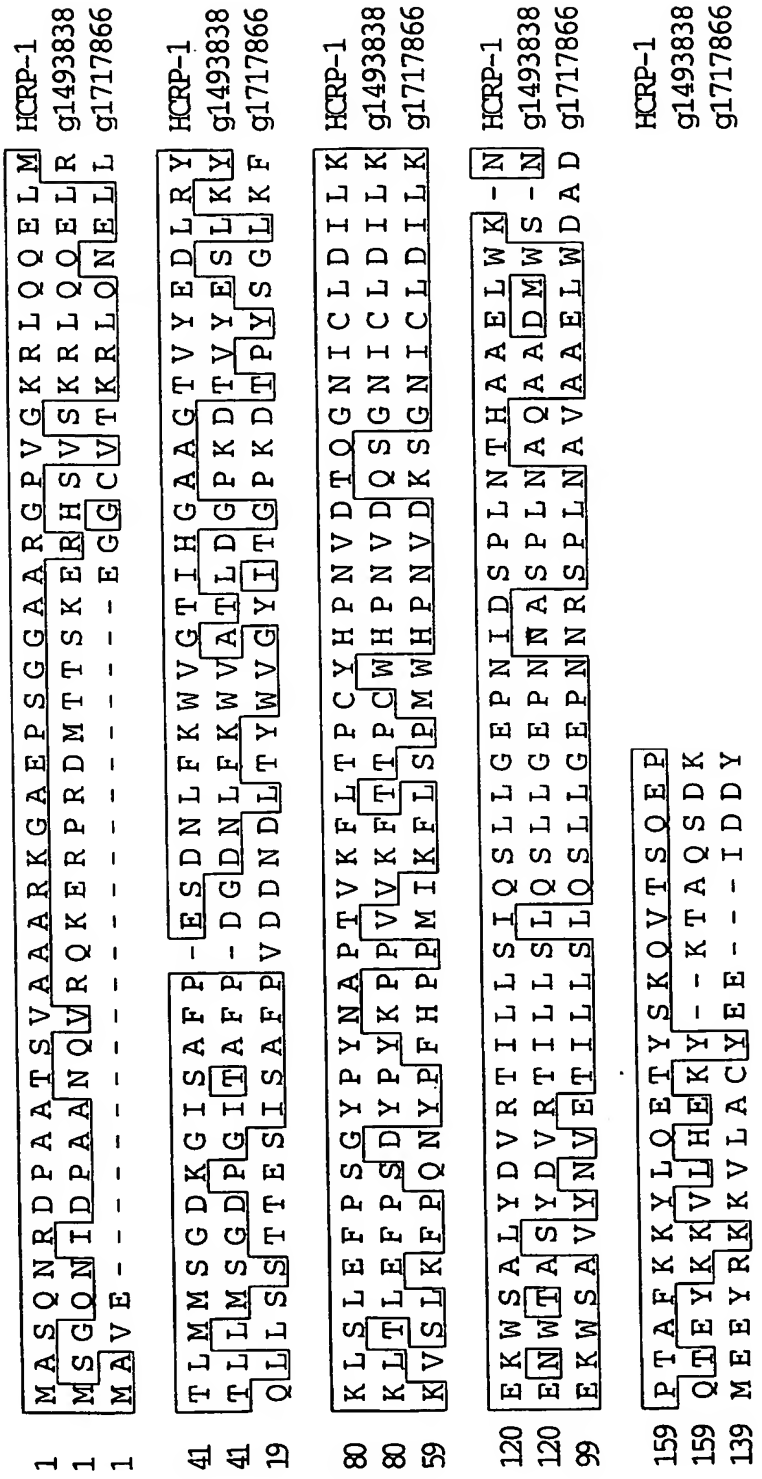
1 MASQNRDPAATSVAAARKGAEPSGGAARGPVGKRRLQQELM HCRP-1
1 MSGQNIDPAANQVRQKERPRDMTTSKERHSVSKRRLQQELR g1493838
1 MAVE----------E GGCVTKRRLQNELL g1717866
41 TLMMSGDKGISAFP-ESDNLFKWVGTIHGAAGTVYEDLRY HCRP-1
41 TLLMSGDPGITAFP-DGDNLFKWVATLDGPKDTVYESLKY g1493838
19 QLLSSTTESISAFPVDDNDLTYWVGYITGPKDTPYSGLKF g1717866
80 KLSLEFPSGYPYNAPTVKFLTPCYHPNVDTQGNICLDIIK HCRP-1
80 KLTLEFPSDYPPYKPPVVKFTTPCWHPNVDQSGNICLDIIK g1493838
59 KVSLKFPQNYPFHPPMIKFLSPMWHPNVDKSGNICLDILK g1717866
120 EKWSALYDVRTILLSIQSLLGEPNIDSPLNTHAAELWK-N HCRP-1
120 ENWTASYDVRTILLSLQSLLGEPNWASPLNAQAADMWS-N g1493838
99 EKWSAVYNVETILLSLQSLLGEPNNRSPLNAVAAELWD A D g1717866
159 PTAFKKYLQETYSKQVTSQEP HCRP-1
159 QTEYKKVLHEKY--KTAQSDK g1493838
139 MEEYRKKVLACYEE---IDDY g1717866

FIGURE 3

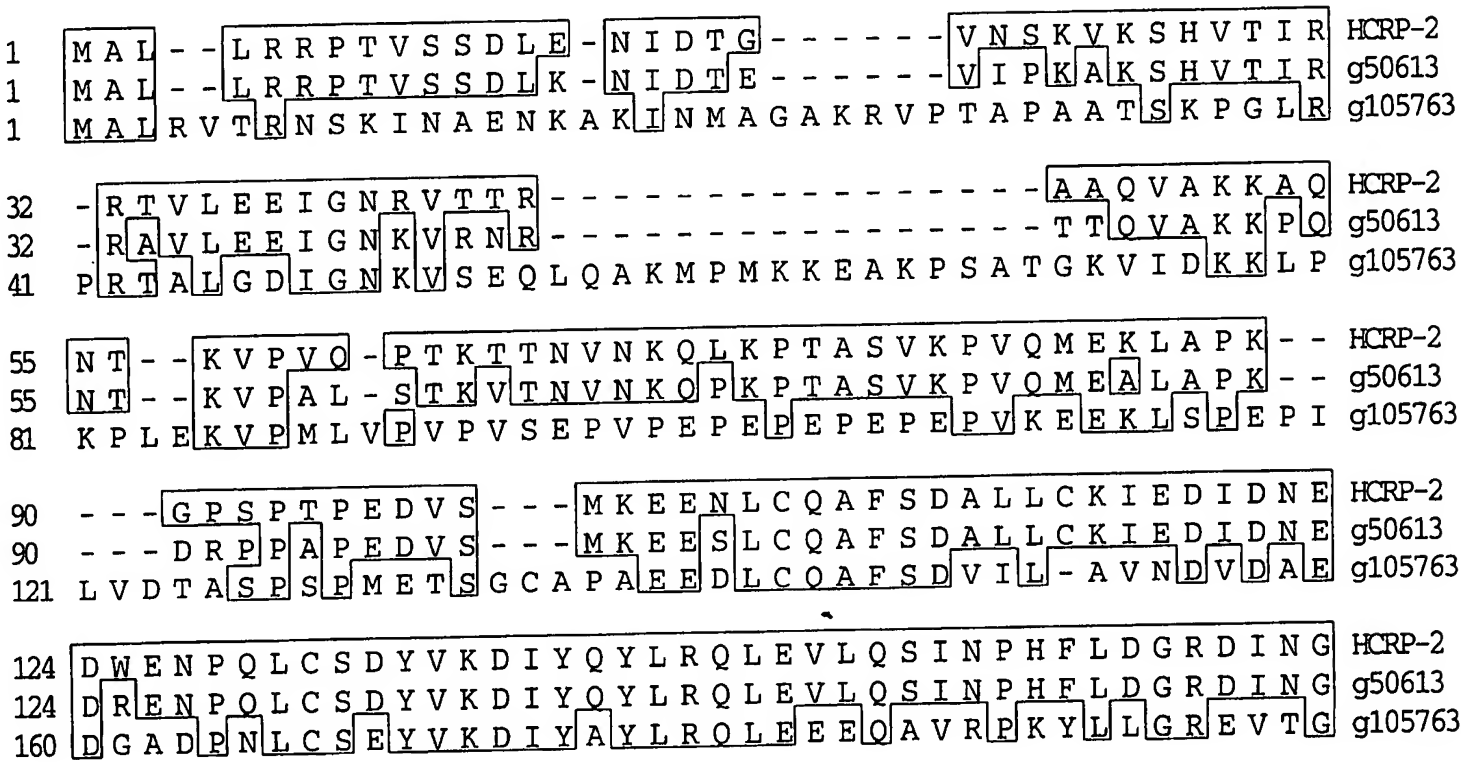
1 MAL--LRRPTVSSDLE-NIDTG------VNSKVKSHVTIR HCRP-2
1 MAL--LRRPTVSSDLK-NIDTE------VIPKAKSHVTIR g50613
1 MALRVTRNSKINAENKAKINMAGAKRVPTAPAATSKPGLR g105763
32 -RTVLEEIGNRVTTR----------AAQVAKKAQ HCRP-2
32 -RAVLEEIGNKVRNR----------TTQVAKKPQ g50613
41 PRTALGDIGNKVSEQLQAKMPMKKEAKPSATGKVIDKKLP g105763
55 NT--KVPVQ-PTKTTNVNKQLKPTASVKPVQMEKLAPK-- HCRP-2
55 NT--KVPAL-STKVTNVNKQPKPTASVKPVQMEALAPK-- g50613
81 KPLEKVPMLVPVPVSEPVPEPEPEPEPEPVKEEKLSPEPI g105763
90 ---GPSPTPEDVS---MKEENLCQAFSDALLCKIEDIDNE HCRP-2
90 ---DRPPAPEDVS---MKEESLCQAFSDALLCKIEDIDNE g50613
121 LVDTASPSPMETSGCAPAEEDLCQAFSDVIL-AVNDVDAE g105763
124 DWENPQLCSDYVKDIYQYLRQLEVLQSINPHFLDGRDING HCRP-2
124 DRENPQLCSDYVKDIYQYLRQLEVLQSINPHFLDGRDING g50613
160 DGADPNLCSEYVKDIYAYLRQLEEEQAVRPKYLLGREVTG g105763

FIGURE 4A

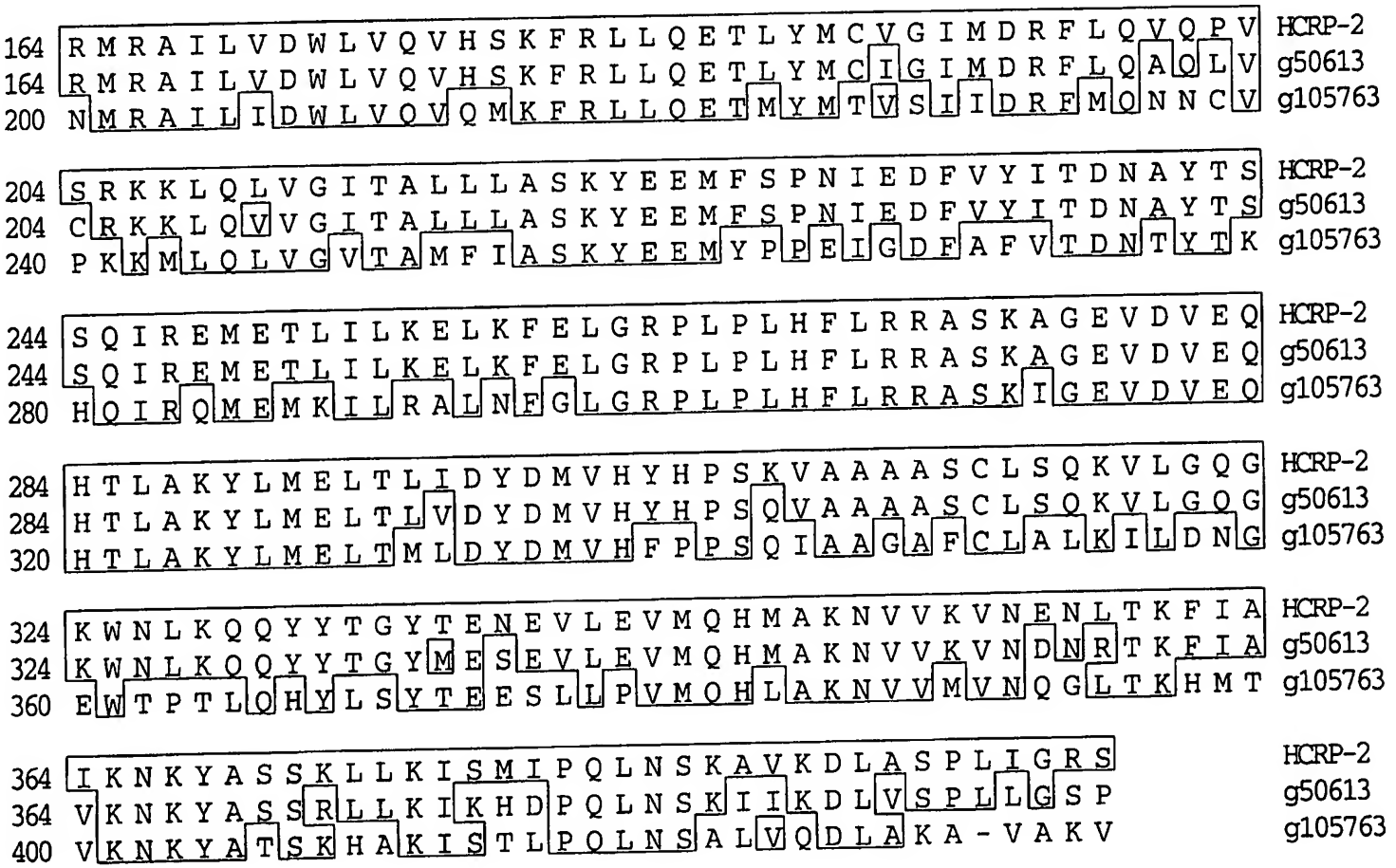
164 R M R A I L V D W L V Q V H S K F R L L Q E T L Y M C V G I M D R F L Q V Q P V HCRP-2
164 R M R A I L V D W L V Q V H S K F R L L Q E T L Y M C I G I M D R F L Q A Q L V g50613
200 N M R A I L I D W L V Q V Q M K F R L L Q E T M Y M T V S I I D R F M Q N N C V g105763
204 S R K K L Q L V G I T A L L L A S K Y E E M F S P N I E D F V Y I T D N A Y T S HCRP-2
204 C R K K L Q V V G I T A L L L A S K Y E E M F S P N I E D F V Y I T D N A Y T S g50613
240 P K K M L Q L V G V T A M F I A S K Y E E M Y P P E I G D F A F V T D N T Y T K g105763
244 S Q I R E M E T L I L K E L K F E L G R P L P L H F L R R A S K A G E V D V E Q HCRP-2
244 S Q I R E M E T L I L K E L K F E L G R P L P L H F L R R A S K A G E V D V E Q g50613
280 H Q I R Q M E M K I L R A L N F G L G R P L P L H F L R R A S K I G E V D V E Q g105763
284 H T L A K Y L M E L T L I D Y D M V H Y H P S K V A A A A S C L S Q K V L G Q G HCRP-2
284 H T L A K Y L M E L T L V D Y D M V H Y H P S Q V A A A A S C L S Q K V L G Q G g50613
320 H T L A K Y L M E L T M L D Y D M V H F P P S Q I A A G A F C L A L K I L D N G g105763
324 K W N L K Q Q Y Y T G Y T E N E V L E V M Q H M A K N V V K V N E N L T K F I A HCRP-2
324 K W N L K Q Q Y Y T G Y M E S E V L E V M Q H M A K N V V K V N D N R T K F I A g50613
360 E W T P T L Q H Y L S Y T E E S L L P V M Q H L A K N V V M V N Q G L T K H M T g105763
364 I K N K Y A S S K L L K I S M I P Q L N S K A V K D L A S P L I G R S HCRP-2
364 V K N K Y A S S R L L K I K H D P Q L N S K I I K D L V S P L L G S P g50613
400 V K N K Y A T S K H A K I S T L P Q L N S A L V Q D L A K A - V A K V g105763

FIGURE 4B

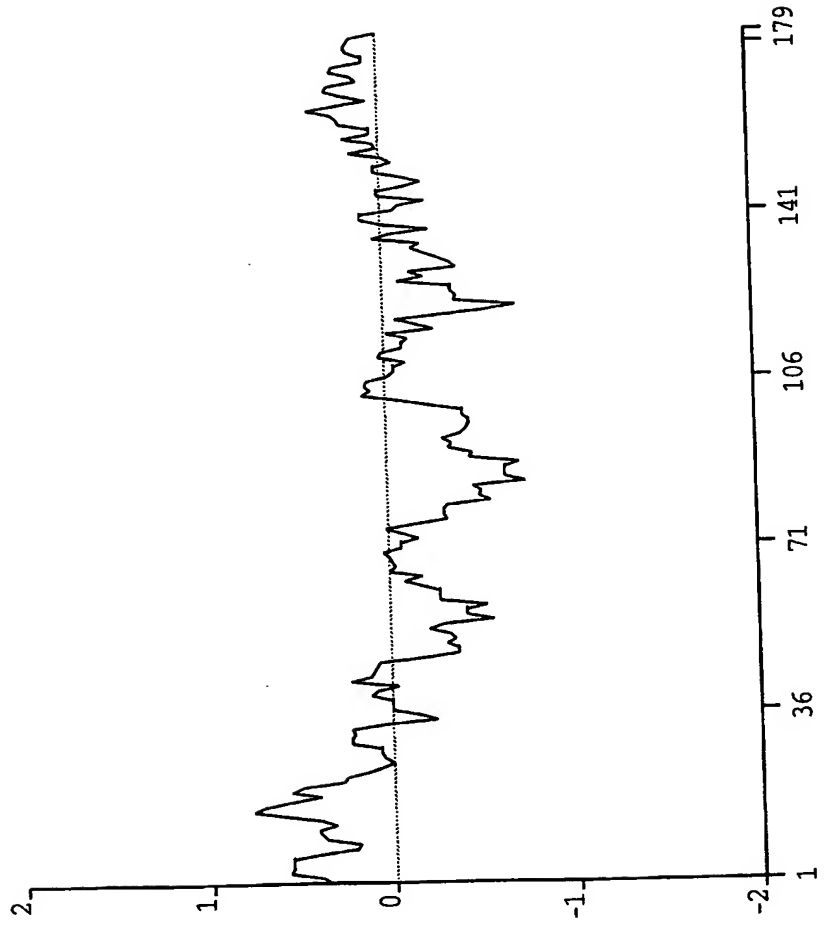

FIGURE 5A

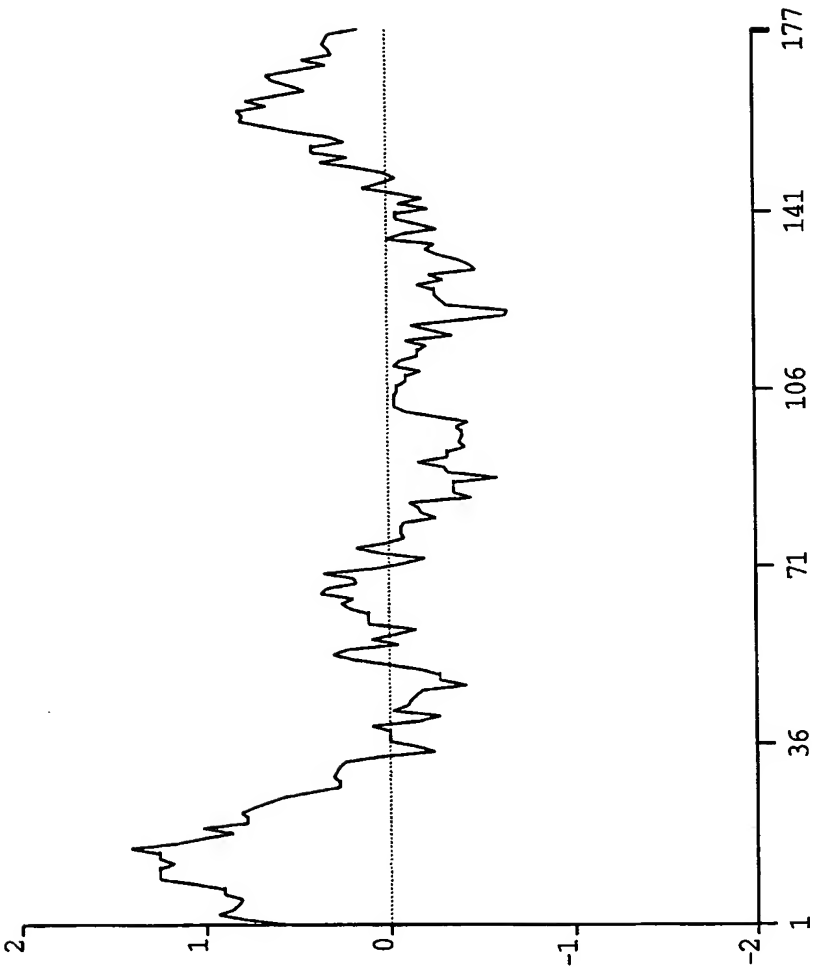

FIGURE 5B

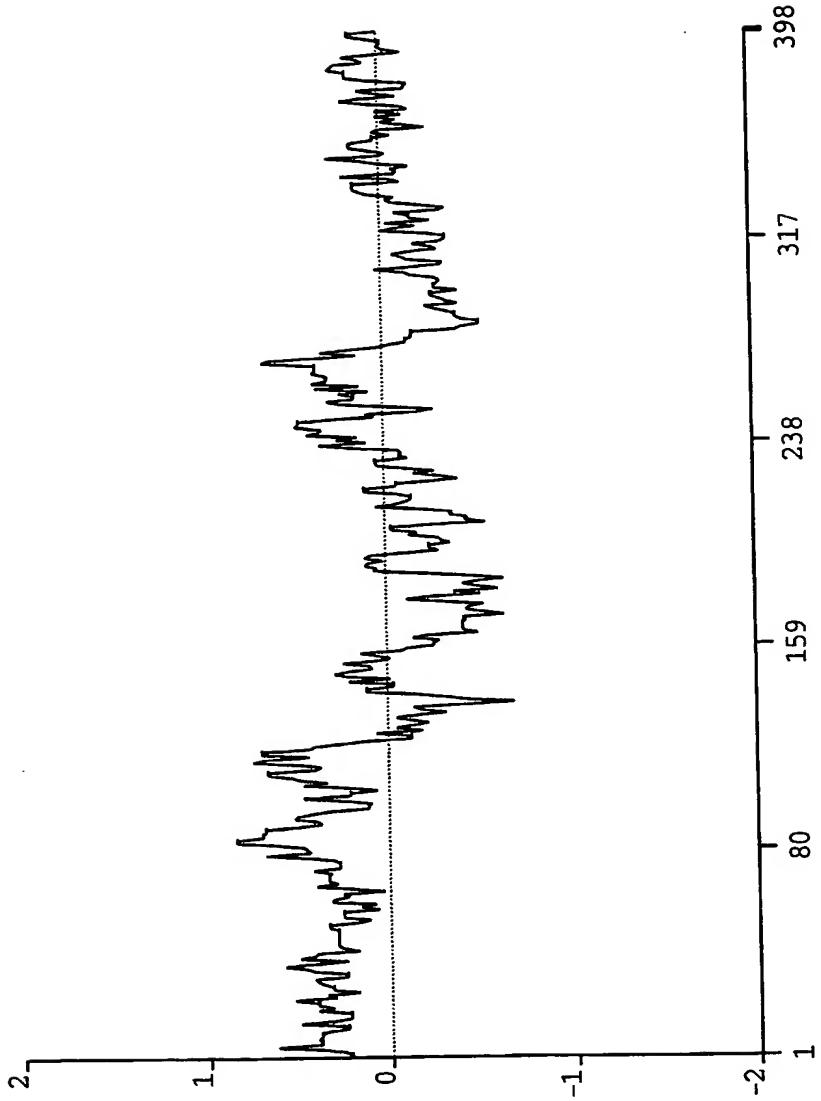

FIGURE 6A

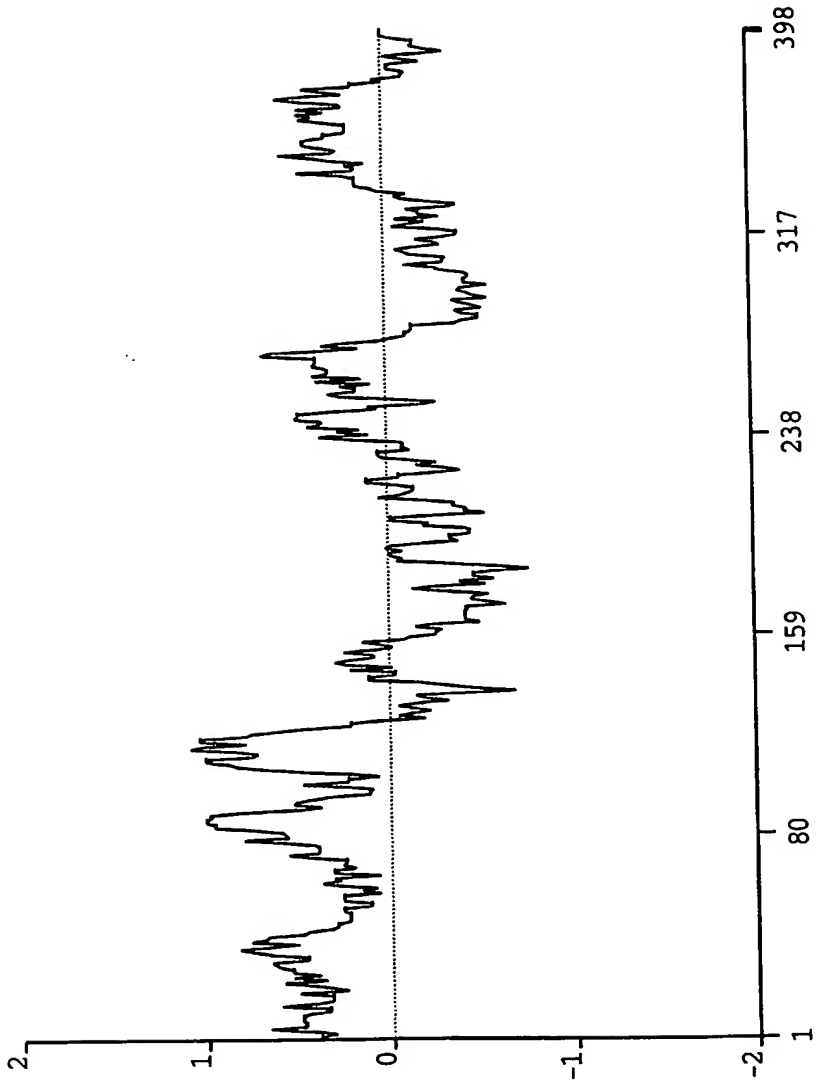

FIGURE 6B

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/52, 15/12, 9/00, C07K 14/47 A61K 38/43, 38/17, C12Q 1/68, C07K 16/18, 16/40

A3

(11) International Publication Number: **WO 98/55510**

(43) International Publication Date: 10 December 1998 (10.12.98)

(21) International Application Number: PCT/US98/11735

(22) International Filing Date: 5 June 1998 (05.06.98)

(30) Priority Data:
08/870,143 5 June 1997 (05.06.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 08/870,143 (CIP)
Filed on 5 June 1997 (05.06.97)

(71) Applicant *(for all designated States except US)*: INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). GUEGLER, Karn, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). COCKS, Benjamin, Graeme [AU/US]; 4292 D. Wilke Way, Palo Alto, CA 94306 (US). SHAH, Purvi [IN/US]; 1608 Queen Charlotte Drive, Sunnyvale, CA 94087 (US).

(74) Agent: BILLINGS, Lucy, J.; Incyte Pharmaceuticals, Inc., 3147 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AT, AU, BR, CA, CH, CN, DE, DK, ES, FI, GB, IL, JP, KR, MX, NO, NZ, RU, SE, SG, US, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(88) Date of publication of the international search report:
18 March 1999 (18.03.99)

(54) Title: HUMAN UBIQUITIN CARRIER PROTEIN E2-LIKE PROTEIN AND CYCLIN B-LIKE PROTEIN

```
1   MASQNRDPAATSVAAARKGAEPSGGAARGPVGKRLQQELM  HCRP-1
1   MSGQNIDPAANQVRQKERPRDMTTSKERHSVSKRLQQELR  g1493838
1   MAVE----------------------EGGCVTKRLQNELL  g1717866

41  TLMMSGDKGISAFP-ESDNLFKWVGTIHGAAGTVYEDLRY  HCRP-1
41  TLLMSGDPGITAFP-DGDNLFKWVATLDGPKDTVYESLKY  g1493838
19  QLLSSTTESISAFPVDDNDLTYWVGYITGPKDTPYSGLKF  g1717866

80  KLSLEFPSGYPYNAPTVKFLTPCYHPNVDTQGNICLDILK  HCRP-1
80  KLTLEFPSDYPYKPPVVKFTTPCWHPNVDQSGNICLDILK  g1493838
59  KVSLKFPQNYPFHPPMIKFLSPMWHPNVDKSGNICLDILK  g1717866

120 EKWSALYDVRTILLSIQSLLGEPNIDSPLNTHAAELWK-N  HCRP-1
120 ENWTASYDVRTILLSLQSLLGEPNNASPLNAQAADMWS-N  g1493838
99  EKWSAVYNVETILLSLQSLLGEPNNRSPLNAVAAELWDAD  g1717866

159 PTAFKKYLQETYSKQVTSQEP                     HCRP-1
159 QTEYKKVLHEKY--KTAQSDK                     g1493838
139 MEEYRKKVLACYEE---IDDY                     g1717866
```

(57) Abstract

The invention provides two human proteins, involved in the cyclin-mediated regulatory pathway; a cyclin B-like protein, designated HCRP-2, and a ubiquitin conjugating enzyme E2-like protein, allegedly specifically involved in the degradation of mitotic cyclins, designated HCRP-1.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INTERNATIONAL SEARCH REPORT

International Application No
PCT/. . 98/11735

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/52 C12N15/12 C12N9/00 C07K14/47 A61K38/43
A61K38/17 C12Q1/68 C07K16/18 C07K16/40

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K A61K C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	TOWNSLEY F M ET AL: "DOMINANT-NEGATIVE CYCLIN-SELECTIVE UBIQUITIN CARRIER PROTEIN E2-C/UBCH10 BLOCKS CELLS IN METAPHASE" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 94, no. 6, 18 March 1997, pages 2362-2367, XP002039885 see the whole document ---	1-11,14, 17,18
P,X	WO 97 37027 A (HARVARD COLLEGE) 9 October 1997 see the whole document -----	1-18

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
3 September 1998	0 3. 12. 98

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Smalt, R

Form PCT/ISA/210 (second sheet) (July 1992)

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-18

A ubiquitin conjugating enzyme E2-like human protein as defined in seq.ID 2 or fragments thereof, nucleic acids which encode them, or are complementary or hybridizing thereto, (pharmaceutical) compositions comprising said protein or nucleic acid, expression vector comprising said nucleic acid, host cell comprising said vector and method of producing the protein using said host cell. Also an antibody against said protein, agonists of the protein, and method for detecting the nucleic acid trhough hybridization and/or PCR.

2. Claims: 19-37

A cyclin B-like human protein as defined in seq.ID 2 or fragments thereof, nucleic acids which encode them, or are complementary or hybridizing thereto, (pharmaceutical) compositions comprising said protein or nucleic acid, expression vector comprising said nucleic acid, host cell comprising said vector and method of producing the protein using said host cell. Also an antibody against said protein, (ant)agonists for the protein, and method for detecting the nucleic acid trhough hybridization and/or PCR.

REMARK:
Although claims 15 and 16 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Further defect(s) under Article 17(2)(a):

Claims Nos: 14 (partially)

The search for the agonist could only be performed partially due to insufficient characterisation of the compound.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 98/11735

Box III TEXT OF THE ABSTRACT (Continuation of item 5 of the first sheet)

The invention provides two human proteins, involved in the cyclin-mediated regulatory pathway; a cyclin B-like protein, designated HCRP-2, and a ubiquitin conjugating enzyme E2-like protein, allegedly specifically involved in the degradation of mitotic cyclins, designated HCRP-1.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 98/11735

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
Please see Further Information sheet enclosed.

2. [X] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
Please see Further Information sheet enclosed.

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Claims 1-18

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 98/11735

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9737027 A	09-10-97	AU 2600697 A	22-10-97